

## (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
14 June 2001 (14.06.2001)

PCT

(10) International Publication Number  
**WO 01/42285 A2**

(51) International Patent Classification<sup>7</sup>: C07K 14/00

(21) International Application Number: PCT/US00/32990

(22) International Filing Date: 5 December 2000 (05.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/172,852 10 December 1999 (10.12.1999) US  
60/172,354 16 December 1999 (16.12.1999) US

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). BURFORD, Neil [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). AZIMZAI, Yalda [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way, #1, Menlo Park, CA 94025 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). LU, Dyung, Aina,

M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). SHAH, Purvi [IN/US]; 859 Salt Lake Drive, San Jose, CA 95133 (US). AU-YOUNG, Janice [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES

(57) Abstract: The invention provides human extracellular matrix and cell adhesion molecules (XMAD) and polynucleotides which identify and encode XMAD. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of XMAD.

WO 01/42285 A2

WO 01/42285

PCT/US00/32990

## EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of extracellular matrix and cell adhesion molecules and to the use of these sequences in the diagnosis, treatment, and prevention of genetic, autoimmune/inflammation, and cell proliferative disorders, including cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of extracellular matrix and cell adhesion molecules.

### BACKGROUND OF THE INVENTION

#### Extracellular Matrix Proteins

The extracellular matrix (ECM) is a complex network of glycoproteins, polysaccharides, proteoglycans, and other macromolecules that are secreted from the cell into the extracellular space. The ECM remains in close association with the cell surface and provides a supportive meshwork that profoundly influences cell shape, motility, strength, flexibility, and adhesion. In fact, adhesion of a cell to its surrounding matrix is required for cell survival except in the case of metastatic tumor cells, which have overcome the need for cell-ECM anchorage. This phenomenon suggests that the ECM plays a critical role in the molecular mechanisms of growth control and metastasis. (Reviewed in Ruoslahti, E. (1996) Sci. Am. 275:72-77.) Furthermore, the ECM determines the structure and physical properties of connective tissue and is particularly important for morphogenesis and other processes associated with embryonic development and pattern formation.

The collagens comprise a family of ECM proteins that provide structure to bone, teeth, skin, ligaments, tendons, cartilage, blood vessels, and basement membranes. Multiple collagen proteins have been identified. Three collagen molecules fold together in a triple helix stabilized by interchain disulfide bonds. Bundles of these triple helices then associate to form fibrils.

Elastin and related proteins confer elasticity to tissues such as skin, blood vessels, and lungs. Elastin is a highly hydrophobic protein of about 750 amino acids that is rich in proline and glycine residues. Elastin molecules are highly cross-linked, forming an extensive extracellular network of fibers and sheets. Elastin fibers are surrounded by a sheath of microfibrils which are composed of a number of glycoproteins, including fibrillin.

Fibronectin is a large ECM glycoprotein found in all vertebrates. Fibronectin exists as a dimer of two subunits, each containing about 2,500 amino acids. Each subunit folds into a rod-like structure containing multiple domains. The domains each contain multiple repeated modules, the most common of which is the type III fibronectin repeat. The type III fibronectin repeat is about 90 amino acids in

WO 01/42285

PCT/US00/32990

length and is also found in other ECM proteins and in some plasma membrane and cytoplasmic proteins. Furthermore, some type III fibronectin repeats contain a characteristic tripeptide consisting of Arginine-Glycine-Aspartic acid (RGD). The RGD sequence is recognized by the integrin family of cell surface receptors and is also found in other ECM proteins. (Reviewed in Alberts, supra, pp. 986-987.)

5 Laminin is a major glycoprotein component of the basal lamina which underlies and supports epithelial cell sheets. Laminin is one of the first ECM proteins synthesized in the developing embryo. Laminin is an 850 kilodalton protein composed of three polypeptide chains joined in the shape of a cross by disulfide bonds. Laminin is especially important for angiogenesis and, in particular, for guiding the formation of capillaries. (Reviewed in Alberts, B., et al. (1994) Molecular Biology of the  
10 Cell, Garland Publishing, New York, NY, pp. 990-991.)

Many proteinaceous ECM components are proteoglycans. Proteoglycans are composed of unbranched polysaccharide chains (glycosaminoglycans) attached to protein cores. Common proteoglycans include aggrecan, betaglycan, decorin, perlecan, serglycin, and syndecan-1. Some of these molecules not only provide mechanical support, but also bind to extracellular signaling molecules,  
15 such as fibroblast growth factor and transforming growth factor  $\beta$ , suggesting a role for proteoglycans in cell-cell communication. (Reviewed in Alberts, supra, pp. 973-978.)

Dentin phosphoryn (DPP) is a major component of the dentin ECM. DPP is a proteoglycan that is synthesized and expressed by odontoblasts (Gu, K., et al. (1998) *Eur. J. Oral Sci.* 106:1043-1047). DPP is believed to nucleate or modulate the formation of hydroxyapatite crystals.

20 Mucins are highly glycosylated glycoproteins that are the major structural component of the mucus gel. The physiological functions of mucins are cytoprotection, mechanical protection, maintenance of viscosity in secretions, and cellular recognition. MUC6 is a human gastric mucin that is also found in gall bladder, pancreas, seminal vesicles, and female reproductive tract (Toribara, N.W., et al. (1997) *J. Biol. Chem.* 272:16398-16403). The MUC6 gene has been mapped to human  
25 chromosome 11 (Toribara, N.W., et al. (1993) *J. Biol. Chem.* 268:5879-5885). Hemomucin is a novel Drosophila surface mucin that may be involved in the induction of antibacterial effector molecules (Theopold, U., et al. (1996) *J. Biol. Chem.* 271:12708-12715).

Extracellular matrix proteins may regulate cellular protein activity in a variety of ways. Reversible protein phosphorylation is the primary method for regulating protein activity in eukaryotic  
30 cells. In general, proteins are activated by phosphorylation in response to extracellular signals such as hormones, neurotransmitters, and growth and differentiation factors. The activated proteins initiate the cell's intracellular response by way of intracellular signaling pathways and second messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens, that regulate protein phosphorylation.

WO 01/42285

PCT/US00/32990

### Adhesion-Associated Proteins

The surface of a cell is rich in transmembrane proteoglycans, glycoproteins, glycolipids, and receptors. These macromolecules mediate adhesion with other cells and with components of the ECM. The interaction of the cell with its surroundings profoundly influences cell shape, strength, flexibility, motility, and adhesion. These dynamic properties are intimately associated with signal transduction pathways controlling cell proliferation and differentiation, tissue construction, and embryonic development.

Cadherins comprise a family of calcium-dependent glycoproteins that function in mediating cell-cell adhesion in virtually all solid tissues of multicellular organisms. These proteins share multiple repeats of a cadherin-specific motif, and the repeats form the folding units of the cadherin ECM. Cadherin molecules cooperate to form focal contacts, or adhesion plaques, between adjacent epithelial cells. The cadherin family includes the classical cadherins and protocadherins. Classical cadherins include the E-cadherin, N-cadherin, and P-cadherin subfamilies. E-cadherin is present on many types of epithelial cells and is especially important for embryonic development. P-cadherin is present on cells of the placenta and epidermis. Recent studies report that protocadherins are involved in a variety of cell-cell interactions (Suzuki, S. T. (1996) *J. Cell Sci.* 109:2609-2611). The intracellular anchorage of cadherins is regulated by their dynamic association with catenins, a family of cytoplasmic signal transduction proteins associated with the actin cytoskeleton. The anchorage of cadherins to the actin cytoskeleton appears to be regulated by protein tyrosine phosphorylation, and the cadherins are the target of phosphorylation-induced junctional disassembly (Aberle, H., et al. (1996) *J. Cell. Biochem.* 61:514-523).

Integrins are ubiquitous transmembrane adhesion molecules that link the ECM to the internal cytoskeleton. Integrins are composed of two noncovalently associated transmembrane glycoprotein subunits called  $\alpha$  and  $\beta$ . Integrins function as receptors that play a role in signal transduction. For example, binding of integrin to its extracellular ligand may stimulate changes in intracellular calcium levels or protein kinase activity (Sjaastad, M.D. and Nelson, W.J. (1997) *BioEssays* 19:47-55).

Lectins comprise a ubiquitous family of extracellular glycoproteins which bind cell surface carbohydrates specifically and reversibly, resulting in the agglutination of cells. (Reviewed in Drickamer, K. and Taylor, M.E. (1993) *Annu. Rev. Cell Biol.* 9:237-264.) This function is particularly important for activation of the immune response. Lectins mediate the agglutination and mitogenic stimulation of lymphocytes at sites of inflammation (Lasky, L.A. (1991) *J. Cell. Biochem.* 45:139-146; Paietta, E., et al. (1989) *J. Immunol.* 143:2850-2857). C-type lectin domains are found in a variety of proteins, including selectins and lecticans. Lecticans are a family of chondroitin sulfate proteoglycans that include aggrecan, versican, neurocan, and brevican. All C-type lectin proteins are



WO 01/42285

PCT/US00/32990

involved in protein-protein interactions (Aspberg, A., et al. (1997) Proc. Natl. Acad. Sci. USA 94:10116-10121). A novel macrophage-restricted C-type lectin protein has been cloned from mouse tissue. It is a type II transmembrane protein with one extracellular C-type lectin domain (Balch, S.G., et al. (1998) J. Biol. Chem. 273:18656-18664).

5 Toposome is a cell-adhesion glycoprotein isolated from mesenchyme-blastula embryos. Toposome precursors including vitellogenin promote cell adhesion of dissociated blastula cells.

LRRs are sequence motifs, approximately 22-28 amino acids in length, found in proteins with a large variety of functions and cellular locations. Proteins containing LRRs are all thought to be involved in protein-protein interactions. The crystal structure of LRRs has been studied and found to  
10 correspond to beta-alpha structural units. These structural units form a parallel beta sheet with one surface exposed to solvent. In this way an LRR-containing protein acquires a nonglobular shape (Kobe, B. and Deisenhofer, J. (1994) Trends Biochem. Sci. 19:415-421). There is evidence to suggest LRRs function in signal transduction and cellular adhesion as well as in protein-protein interactions (Gay, N.J., et al. (1991) FEBS Lett. 29:87-91).

15 Various proteins such as those encoded by the Drosophila armadillo gene and the human APC gene contain amino acid repeats that interact with  $\beta$ -catenins. The armadillo gene is required for pattern formation within the embryonic segments and imaginal discs and is highly conserved. It is 63% identical to a human protein, plakoglobin, which is involved in adhesive junctions joining epithelial and other cells (Peifer, M. and Wieschaus, E. (1990) Cell 63:1167-1176). APC gene mutations appear to  
20 initiate inherited forms of human colorectal cancer and sporadic forms of colorectal and gastric cancer (Rubinfeld, B., et al. (1993) Science 262:1731-1734). The fact that the protein encoded by APC interacts with catenin suggests a link between tumor initiation and cell adhesion (Su, L.K., et al. (1993) Science 262:1734-1737).

SH3 is a 60-70 amino acid motif found in a variety of signal transduction and cytoskeletal  
25 proteins. The SH3 domain is involved in mediating protein-protein interactions. Evidence suggests that the SH3 domains recognize a family of related domains or proteins in a variety of different tissues and species. One novel SH3 domain-containing protein is the 52 kilodalton focal adhesion protein (FAP52 or p52). FAP52 is localized to focal adhesions, specialized membrane domains in cultured cells that mediate the attachment of cells to the growth substratum and ECM. Focal adhesions consist of  
30 structural proteins, integrins, regulatory molecules, and signaling molecules and are involved in cell signaling. FAP52 may form part of this multimolecular complex that comprises focal adhesion sites (Merilainen, J., et al. (1997) J. Biol. Chem. 272:23278-23284).

The discovery of new extracellular matrix and cell adhesion molecules and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the

WO 01/42285

PCT/US00/32990

diagnosis, prevention, and treatment of genetic, autoimmune/inflammation, and cell proliferative disorders, including cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of extracellular matrix and cell adhesion molecules.

5

### SUMMARY OF THE INVENTION

The invention features purified polypeptides, extracellular matrix and cell adhesion molecules, referred to collectively as "XMAD" and individually as "XMAD-1," "XMAD-2," "XMAD-3," "XMAD-4," "XMAD-5," "XMAD-6," "XMAD-7," "XMAD-8," "XMAD-9," "XMAD-10," "XMAD-11," "XMAD-12," "XMAD-13," "XMAD-14," "XMAD-15," "XMAD-16," "XMAD-17," "XMAD-18," "XMAD-19," "XMAD-20," and "XMAD-21." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-21.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-21. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:22-42.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group

WO 01/42285

PCT/US00/32990

consisting of SEQ ID NO:1-21. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid  
5 sequence selected from the group consisting of a) an amino acid sequence selected from the group  
consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90%  
sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c)  
a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID  
NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group  
10 consisting of SEQ ID NO:1-21. The method comprises a) culturing a cell under conditions suitable for  
expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide  
comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b)  
recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a  
15 polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid  
sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid  
sequence having at least 90% sequence identity to an amino acid sequence selected from the group  
consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected  
from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid  
20 sequence selected from the group consisting of SEQ ID NO:1-21.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence  
selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of  
SEQ ID NO:22-42, b) a naturally occurring polynucleotide sequence having at least 90% sequence  
identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a  
25 polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e)  
an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous  
nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample,  
said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence  
30 selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of  
SEQ ID NO:22-42, b) a naturally occurring polynucleotide sequence having at least 90% sequence  
identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a  
polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e)  
an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising

WO 01/42285

PCT/US00/32990

at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and  
5 optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of  
10 SEQ ID NO:22-42, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or  
15 absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence  
20 having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of  
25 SEQ ID NO:1-21. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional XMAP, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an  
30 amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a)

WO 01/42285

PCT/US00/32990

exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional XMAD, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of, a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional XMAD, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The

WO 01/42285

PCT/US00/32990

method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:22-42, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

#### BRIEF DESCRIPTION OF THE TABLES

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs),

WO 01/42285

PCT/US00/32990

clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding XMAD.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods, algorithms, and searchable databases used for analysis of XMAD.

5 Table 3 shows selected fragments of each nucleic acid sequence; the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding XMAD were isolated.

10 Table 5 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

### DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood  
15 that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an,"  
20 and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings  
25 as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in  
30 connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

### DEFINITIONS

"XMAD" refers to the amino acid sequences of substantially purified XMAD obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and

WO 01/42285

PCT/US00/32990

human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of XMAD. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of XMAD either by directly interacting with XMAD or by acting on components of the biological pathway in which XMAD participates.

An "allelic variant" is an alternative form of the gene encoding XMAD. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding XMAD include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as XMAD or a polypeptide with at least one functional characteristic of XMAD. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding XMAD, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding XMAD. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent XMAD. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of XMAD is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known



WO 01/42285

PCT/US00/32990

in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of XMAD. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of XMAD either by  
5 directly interacting with XMAD or by acting on components of the biological pathway in which XMAD participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind XMAD polypeptides can be prepared using intact polypeptides or using fragments  
10 containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

15 The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to  
20 elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified  
25 sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or  
30 translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic XMAD, or of any oligopeptide thereof,

WO 01/42285

PCT/US00/32990

to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement,

5 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution.

Compositions comprising polynucleotide sequences encoding XMAD or fragments of XMAD may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
30	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
35	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu

WO 01/42285

PCT/US00/32990

5	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

---

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

A "fragment" is a unique portion of XMAD or the polynucleotide encoding XMAD which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:22-42 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:22-42, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:22-42 is useful, for

WO 01/42285

PCT/US00/32990

example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:22-42 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:22-42 and the region of SEQ ID NO:22-42 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

5 A fragment of SEQ ID NO:1-21 is encoded by a fragment of SEQ ID NO:22-42. A fragment of SEQ ID NO:1-21 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-21. For example, a fragment of SEQ ID NO:1-21 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-21. The precise length of a fragment of SEQ ID NO:1-21 and the region of SEQ ID NO:1-21 to which the fragment  
10 corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full-length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full-length" polynucleotide sequence encodes a "full-length" polypeptide sequence.

15 "Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in  
20 the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular  
25 biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between  
30 aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at  
35 <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis

WO 01/42285

PCT/US00/32990

programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62  
 Reward for match: 1  
 Penalty for mismatch: -2  
 Open Gap: 5 and Extension Gap: 2 penalties  
 Gap x drop-off: 50  
 Expect: 10  
 Word Size: 11  
 Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters

WO 01/42285

PCT/US00/32990

of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (Apr-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

10     *Matrix: BLOSUM62*  
        *Open Gap: 11 and Extension Gap: 1 penalties*  
        *Gap x drop-off: 50*  
        *Expect: 10*  
        *Word Size: 3*  
 15     *Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e.,

WO 01/42285

PCT/US00/32990

binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive  
5 annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic  
10 strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance,  
20 sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is  
25 strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_{ot}$  or  $R_{ot}$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g.,  
30 paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune

WO 01/42285

PCT/US00/32990

disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

5 An "immunogenic fragment" is a polypeptide or oligopeptide fragment of XMAD which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of XMAD which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

10 The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of XMAD. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of XMAD.

15 The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

20 "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

25 "Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

30 "Post-translational modification" of an XMAD may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of XMAD.

35 "Probe" refers to nucleic acid sequences encoding XMAD, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical



WO 01/42285

PCT/US00/32990

labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols In Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that

WO 01/42285

PCT/US00/32990

hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing  
5 primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence.  
10 This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence.  
15 Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated  
20 regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent,  
25 chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose  
30 instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding XMAD, or fragments thereof, or XMAD itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

WO 01/42285

PCT/US00/32990

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The

WO 01/42285

PCT/US00/32990

transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants, and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook, J. et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

The invention is based on the discovery of new human extracellular matrix and cell adhesion molecules (XMAD), the polynucleotides encoding XMAD, and the use of these compositions for the diagnosis, treatment, or prevention of genetic, autoimmune/inflammation, and cell proliferative disorders, including cancer.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding XMAD. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide

WO 01/42285

PCT/US00/32990

and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each XMAD were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries.

- 5 In some cases, GenBank sequence identifiers are also shown in column 5. The Incyte clones and GenBank cDNA sequences, where indicated, in column 5 were used to assemble the consensus nucleotide sequence of each XMAD and are useful as fragments in hybridization technologies.

- The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each  
10 polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows homologous sequences as identified by BLAST analysis along with relevant citations, all of which are expressly incorporated by reference herein in their entirety; and column 7 shows analytical methods and in some cases, searchable databases to which the analytical methods were applied. The  
15 methods of column 7 were used to characterize each polypeptide through sequence homology and protein motifs.

- The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding XMAD. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists fragments of the nucleotide sequences of column 1. These fragments are  
20 useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:22-42 and to distinguish between SEQ ID NO:22-42 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue categories which express XMAD as a fraction of total tissues expressing XMAD. Column 4 lists diseases, disorders, or conditions associated with those tissues expressing XMAD as a fraction of total  
25 tissues expressing XMAD. Column 5 lists the vectors used to subclone each cDNA library.

- The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding XMAD were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

- 30 SEQ ID NO:18 maps to chromosome 22 within the interval from the P terminus to 19.5 centiMorgans.

The invention also encompasses XMAD variants. A preferred XMAD variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the XMAD amino acid sequence, and which contains at least one functional or structural

WO 01/42285

PCT/US00/32990

characteristic of XMAD.

The invention also encompasses polynucleotides which encode XMAD. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:22-42, which encodes XMAD. The polynucleotide sequences of  
 5 SEQ ID NO:22-42, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding XMAD. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least  
 10 about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding XMAD. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:22-42 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:22-42.  
 15 Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of XMAD.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding XMAD, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the  
 20 invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring XMAD, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode XMAD and its variants are generally capable of  
 25 hybridizing to the nucleotide sequence of the naturally occurring XMAD under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding XMAD or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which  
 30 particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding XMAD and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode XMAD and

WO 01/42285

PCT/US00/32990

XMAD derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding XMAD or any fragment thereof.

5 Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:22-42 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in  
10 "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems, Foster City CA), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway  
15 NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or  
20 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

25 The nucleic acid sequences encoding XMAD may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.)  
30 Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al.

WO 01/42285

PCT/US00/32990

(1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060).

5 Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a

10 GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library

15 does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the

20 emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments

25 which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode XMAD may be cloned in recombinant DNA molecules that direct expression of XMAD, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent

30 amino acid sequence may be produced and used to express XMAD.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter XMAD-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic



WO 01/42285

PCT/US00/32990

oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of XMAD, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding XMAD may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, XMAD itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of XMAD, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active XMAD, the nucleotide sequences encoding XMAD or

WO 01/42285

PCT/US00/32990

derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding XMAD. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding XMAD. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding XMAD and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding XMAD and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding XMAD. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington,

WO 01/42285

PCT/US00/32990

J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.)

The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding XMAD. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding XMAD can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding XMAD into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of XMAD are needed, e.g. for the production of antibodies, vectors which direct high level expression of XMAD may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of XMAD. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, *supra*; Bitter, *supra*; and Scorer, *supra*.)

Plant systems may also be used for expression of XMAD. Transcription of sequences encoding XMAD may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, *supra*; Broglie, *supra*; and Winter, *supra*.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., *The McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding XMAD may be ligated into an

WO 01/42285

PCT/US00/32990

adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses XMD in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of XMD in cell lines is preferred. For example, sequences encoding XMD can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*<sup>-</sup> and *aprt*<sup>-</sup> cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

WO 01/42285

PCT/US00/32990

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding XMAD is inserted within a marker gene sequence, transformed cells containing sequences encoding XMAD can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding XMAD under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding XMAD and that express XMAD may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of XMAD using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on XMAD is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding XMAD include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding XMAD, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding XMAD may be cultured under

WO 01/42285

PCT/US00/32990

conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode XMAD may be designed to contain signal sequences which direct  
5 secretion of XMAD through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the  
10 protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding XMAD may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric XMAD protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of XMAD activity. Heterologous protein and  
20 peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins,  
25 respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the XMAD encoding sequence and the heterologous protein sequence, so that XMAD may be cleaved away from the heterologous moiety following purification. Methods for fusion protein  
30 expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled XMAD may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or

WO 01/42285

PCT/US00/32990

SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

XMAD of the present invention or fragments thereof may be used to screen for compounds that specifically bind to XMAD. At least one and up to a plurality of test compounds may be  
 5 screened for specific binding to XMAD. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of XMAD, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):  
 10 Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which XMAD binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express XMAD, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli.  
 15 Cells expressing XMAD or cell membrane fractions which contain XMAD are then contacted with a test compound and binding, stimulation, or inhibition of activity of either XMAD or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example,  
 20 the assay may comprise the steps of combining at least one test compound with XMAD, either in solution or affixed to a solid support, and detecting the binding of XMAD to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a  
 25 solid support.

XMAD of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of XMAD. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for XMAD activity, wherein XMAD is combined with at least one test compound, and the activity of  
 30 XMAD in the presence of a test compound is compared with the activity of XMAD in the absence of the test compound. A change in the activity of XMAD in the presence of the test compound is indicative of a compound that modulates the activity of XMAD. Alternatively, a test compound is combined with an in vitro or cell-free system comprising XMAD under conditions suitable for XMAD activity, and the assay is performed. In either of these assays, a test compound which  
 35 modulates the activity of XMAD may do so indirectly and need not come in direct contact with the

WO 01/42285

PCT/US00/32990

test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding XMAD or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding XMAD may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding XMAD can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding XMAD is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress XMAD, e.g., by secreting XMAD in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

## THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of XMAD and extracellular matrix and cell adhesion molecules. In addition, the expression of XMAD is closely associated with cell proliferation. Therefore, XMAD appears to play



WO 01/42285

PCT/US00/32990

a role in genetic, autoimmune/inflammation, and cell proliferative disorders, including cancer. In the treatment of disorders associated with increased XMAD expression or activity, it is desirable to decrease the expression or activity of XMAD. In the treatment of disorders associated with decreased XMAD expression or activity, it is desirable to increase the expression or activity of XMAD.

5 Therefore, in one embodiment, XMAD or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of XMAD. Examples of such disorders include, but are not limited to, a genetic disorder, such as adrenoleukodystrophy, Alport's syndrome, choroideremia, Duchenne and Becker muscular dystrophy, Down's syndrome, cystic fibrosis, chronic granulomatous disease, Gaucher's disease, 10 Huntington's chorea, Marfan's syndrome, muscular dystrophy, myotonic dystrophy, pycnodysostosis, Refsum's syndrome, retinoblastoma, sickle cell anemia, thalassemia, Werner syndrome, von Willebrand's disease, Wilms' tumor, Zellweger syndrome, peroxisomal acyl-CoA oxidase deficiency, peroxisomal thiolase deficiency, peroxisomal bifunctional protein deficiency, mitochondrial carnitine palmitoyl transferase and carnitine deficiency, mitochondrial very-long-chain acyl-CoA 15 dehydrogenase deficiency, mitochondrial medium-chain acyl-CoA dehydrogenase deficiency, mitochondrial short-chain acyl-CoA dehydrogenase deficiency, mitochondrial electron transport flavoprotein and electron transport flavoprotein:ubiquinone oxidoreductase deficiency, mitochondrial trifunctional protein deficiency, and mitochondrial short-chain 3-hydroxyacyl-CoA dehydrogenase deficiency; an autoimmune/inflammation disorder, such as acquired immunodeficiency syndrome 20 (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, 25 erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, 30 ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers 35 including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma,

WO 01/42285

PCT/US00/32990

and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing XMAD or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of XMAD including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified XMAD in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of XMAD including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of XMAD may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of XMAD including, but not limited to, those listed above.

In a further embodiment, an antagonist of XMAD may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of XMAD. Examples of such disorders include, but are not limited to, those genetic, autoimmune/inflammation, and cell proliferative disorders, including cancer, described above. In one aspect, an antibody which specifically binds XMAD may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express XMAD.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding XMAD may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of XMAD including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of XMAD may be produced using methods which are generally known in the art. In particular, purified XMAD may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind XMAD. Antibodies to XMAD may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit

WO 01/42285

PCT/US00/32990

dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with XMAD or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to XMAD have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of XMAD amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to XMAD may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce XMAD-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for XMAD may also be generated.

WO 01/42285

PCT/US00/32990

For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. 5 (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between XMAD and its 10 specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering XMAD epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for XMAD. Affinity is expressed as an association 15 constant, K<sub>a</sub>, which is defined as the molar concentration of XMAD-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K<sub>a</sub> determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple XMAD epitopes, represents the average affinity, or avidity, of the antibodies for XMAD. The K<sub>a</sub> 20 determined for a preparation of monoclonal antibodies, which are monospecific for a particular XMAD epitope, represents a true measure of affinity. High-affinity antibody preparations with K<sub>a</sub> ranging from about 10<sup>9</sup> to 10<sup>12</sup> L/mole are preferred for use in immunoassays in which the XMAD-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K<sub>a</sub> ranging from about 10<sup>6</sup> to 10<sup>7</sup> L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of XMAD, preferably in active form, from the antibody (Catty, D. 25 (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg 30 specific antibody/ml, is generally employed in procedures requiring precipitation of XMAD-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Colligan et al., supra.)

In another embodiment of the invention, the polynucleotides encoding XMAD, or any fragment

WO 01/42285

PCT/US00/32990

or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding XMAD. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding XMAD. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding XMAD may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA.* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in XMAD expression or regulation causes disease, the expression of

WO 01/42285

PCT/US00/32990

XMAD from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in XMAD are treated by constructing mammalian expression vectors encoding XMAD and introducing these vectors by mechanical means into XMAD-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of XMAD include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). XMAD may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Gossen, M. et al. (1995) *Science* 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) *Curr. Opin. Biotechnol.* 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding XMAD from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to XMAD expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding XMAD under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive

WO 01/42285

PCT/US00/32990

element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding XMAD to cells which have one or more genetic abnormalities with respect to the expression of XMAD. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544; and Verma, I.M. and N. Somia (1997) Nature 389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding XMAD to target cells which have one or more genetic abnormalities with respect to the expression of XMAD. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing XMAD to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S.

WO 01/42285

PCT/US00/32990

Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding XMAD to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full-length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for XMAD into the alphavirus genome in place of the capsid-coding region results in the production of a large number of XMAD-coding RNAs and the synthesis of high levels of XMAD in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of XMAD into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have



WO 01/42285

PCT/US00/32990

been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

5        Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding XMA.

10        Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of  
15        candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis.

20        Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding XMA. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

25        RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and  
30        wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding XMA. Compounds which may be effective in altering expression of a specific polynucleotide may include,  
35        but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming

WO 01/42285

PCT/US00/32990

oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased  
5 XMAD expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding XMAD may be therapeutically useful, and in the treatment of disorders associated with decreased XMAD expression or activity, a compound which specifically promotes expression of the polynucleotide encoding XMAD may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in  
10 altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a  
15 library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding XMAD is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding XMAD are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is  
20 detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding XMAD. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of  
25 the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a  
30 combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for  
35 use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken

WO 01/42285

PCT/US00/32990

from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

5 Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient.

10 Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of XMAD, antibodies to XMAD, and mimetics, agonists, antagonists, or inhibitors of XMAD.

The compositions utilized in this invention may be administered by any number of routes  
15 including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case  
20 of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle  
25 injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of  
30 macromolecules comprising XMAD or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, XMAD or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et

WO 01/42285

PCT/US00/32990

al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example XMAD or fragments thereof, antibodies of XMAD, and agonists, antagonists or inhibitors of XMAD, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind XMAD may be used for the diagnosis of disorders characterized by expression of XMAD, or in assays to monitor patients being

WO 01/42285

PCT/US00/32990

treated with XMAD or agonists, antagonists, or inhibitors of XMAD. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for XMAD include methods which utilize the antibody and a label to detect XMAD in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring XMAD, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of XMAD expression. Normal or standard values for XMAD expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to XMAD under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of XMAD expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding XMAD may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of XMAD may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of XMAD, and to monitor regulation of XMAD levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding XMAD or closely related molecules may be used to identify nucleic acid sequences which encode XMAD. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding XMAD, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the XMAD encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:22-42 or from genomic sequences including promoters, enhancers, and introns of the XMAD gene.

Means for producing specific hybridization probes for DNAs encoding XMAD include the cloning of polynucleotide sequences encoding XMAD or XMAD derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may

WO 01/42285

PCT/US00/32990

be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

5 Polynucleotide sequences encoding XMAD may be used for the diagnosis of disorders associated with expression of XMAD. Examples of such disorders include, but are not limited to, a genetic disorder, such as adrenoleukodystrophy, Alport's syndrome, choroideremia, Duchenne and Becker muscular dystrophy, Down's syndrome, cystic fibrosis, chronic granulomatous disease, Gaucher's disease, Huntington's chorea, Marfan's syndrome, muscular dystrophy, myotonic  
10 dystrophy, pycnodysostosis, Refsum's syndrome, retinoblastoma, sickle cell anemia, thalassemia, Werner syndrome, von Willebrand's disease, Wilms' tumor, Zellweger syndrome, peroxisomal acyl-CoA oxidase deficiency, peroxisomal thiolase deficiency, peroxisomal bifunctional protein deficiency, mitochondrial carnitine palmitoyl transferase and carnitine deficiency, mitochondrial very-long-chain acyl-CoA dehydrogenase deficiency, mitochondrial medium-chain acyl-CoA  
15 dehydrogenase deficiency, mitochondrial short-chain acyl-CoA dehydrogenase deficiency, mitochondrial electron transport flavoprotein and electron transport flavoprotein:ubiquinone oxidoreductase deficiency, mitochondrial trifunctional protein deficiency, and mitochondrial short-chain 3-hydroxyacyl-CoA dehydrogenase deficiency; an autoimmune/inflammation disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress  
20 syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis,  
25 glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner  
30 syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma,  
35 leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of

WO 01/42285

PCT/US00/32990

the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding XMAD may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered XMAD expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding XMAD may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding XMAD may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding XMAD in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of XMAD, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding XMAD, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual

WO 01/42285

PCT/US00/32990

clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding XMAD  
 5 may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding XMAD, or a fragment of a polynucleotide complementary to the polynucleotide encoding XMAD, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of  
 10 closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding XMAD may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation  
 15 polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding XMAD are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable  
 20 using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-  
 25 based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of XMAD include radiolabeling or  
 30 biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid



WO 01/42285

PCT/US00/32990

quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described in Seilhamer, J.J. et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, incorporated herein by reference. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, antibodies specific for XMAD, or XMAD or fragments thereof may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed

WO 01/42285

PCT/US00/32990

molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties.

5 These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data

10 after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is

15 important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the

20 present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present

25 invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and

30 analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent

35 such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is

WO 01/42285

PCT/US00/32990

generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for  
5 example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for XMAD to quantify the  
10 levels of XMAD expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-  
15 reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Sellhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the  
20 analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological  
25 sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

30 In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two

WO 01/42285

PCT/US00/32990

samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. 5 (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding XMAD may be used 10 to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a 15 chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic 20 linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, e.g., Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map 25 data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding XMAD on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as 30 linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized 35 by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences

WO 01/42285

PCT/US00/32990

mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

5 In another embodiment of the invention, XMAD, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between XMAD and the agent being tested may be measured.

10 Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with XMAD, or fragments thereof, and washed. Bound XMAD is then detected by methods well known in the art. Purified XMAD can  
15 also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding XMAD specifically compete with a test compound for binding XMAD.  
20 In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with XMAD.

In additional embodiments, the nucleotide sequences which encode XMAD may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such  
25 properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

30 Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in  
35 particular U.S. Ser. No. 60/172,354, and U.S. Ser. No. 60/172,852, are hereby expressly incorporated

WO 01/42285

PCT/US00/32990

by reference.

## EXAMPLES

### 5 I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl  
10 cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN,  
15 Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP  
20 vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.). Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000  
25 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), pcDNA2.1 plasmid (Invitrogen, Carlsbad CA), or pINCY plasmid (Incyte Genomics, Palo Alto CA). Recombinant  
30 plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

### II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least

WO 01/42285

PCT/US00/32990

one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of  
5 distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using  
10 PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

### III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation  
15 such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).  
20 Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997,  
25 supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VI.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions,  
30 references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between  
35 two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software

WO 01/42285

PCT/US00/32990

Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

5 The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation  
10 using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the  
15 GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and  
20 amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:22-42. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

#### IV. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene  
25 and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is  
30 much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2}) \}}$$



WO 01/42285

PCT/US00/32990

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding XMAD occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

#### V. Chromosomal Mapping of XMAD Encoding Polynucleotides

The cDNA sequences which were used to assemble SEQ ID NO:22-42 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:22-42 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 5). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

The genetic map location of SEQ ID NO:18 is described in The Invention as a range, or interval, of a human chromosome. The map position of an interval, in centiMorgans, is measured

WO 01/42285

PCT/US00/32990

relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

#### VI. Extension of XMAP Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:22-42 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and  $\beta$ -mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100  $\mu$ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5  $\mu$ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the

WO 01/42285

PCT/US00/32990

concentration of DNA. A 5  $\mu$ l to 10  $\mu$ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviII cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, the polynucleotide sequences of SEQ ID NO:22-42 are used to obtain 5' regulatory sequences using the procedure above, along with oligonucleotides designed for such extension, and an appropriate genomic library.

## VII. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:22-42 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10<sup>7</sup> counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human

WO 01/42285

PCT/US00/32990

genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

### VIII. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

### Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is

WO 01/42285

PCT/US00/32990

reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with

5 GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc.

10 (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

#### Microarray Preparation

15 Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia

20 Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and

25 coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic

30 apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in

35 0.2% SDS and distilled water as before.

WO 01/42285

PCT/US00/32990

### Hybridization

Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65 °C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60 °C. The arrays are washed for 10 min at 45 °C in a first wash buffer (1X SSC, 0.1 % SDS), three times for 10 minutes each at 45 °C in a second wash buffer (0.1X SSC), and dried.

### 10 Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

25 The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a

WO 01/42285

PCT/US00/32990

In most expression systems, XMAD is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from XMAD at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified XMAD obtained by these methods can be used directly in the assays shown in Examples XI and XV.

#### XI. Demonstration of XMAD Activity

An assay for XMAD activity measures the disruption of cytoskeletal filament networks upon overexpression of XMAD in cultured cell lines. (Rezniczek, G. A. et al. (1998) J. Cell Biol. 141:209-225.) cDNA encoding XMAD is subcloned into a mammalian expression vector that drives high levels of cDNA expression. This construct is transfected into cultured cells, such as rat kangaroo PtK2 or rat bladder carcinoma 804G cells. Actin filaments and intermediate filaments such as keratin and vimentin are visualized by immunofluorescence microscopy using antibodies and techniques well known in the art. The configuration and abundance of cytoskeletal filaments can be assessed and quantified using confocal imaging techniques. In particular, the bundling and collapse of cytoskeletal filament networks is indicative of XMAD activity.

Alternatively, an assay for XMAD activity measures the amount of cell aggregation induced by overexpression of XMAD. In this assay, cultured cells such as NIH3T3 are transfected with cDNA encoding XMAD contained within a suitable mammalian expression vector under control of a strong promoter. Cotransfection with cDNA encoding a fluorescent marker protein, such as Green Fluorescent Protein (Clontech), is useful for identifying stable transfectants. The amount of cell agglutination, or clumping, associated with transfected cells is compared with that associated with untransfected cells. The amount of cell agglutination is a direct measure of XMAD activity.

Alternatively, protein kinase activity is measured by quantifying the phosphorylation of a protein substrate by XMAD in the presence of gamma-labeled  $^{32}\text{P}$ -ATP. XMAD is incubated with the protein substrate,  $^{32}\text{P}$ -ATP, and an appropriate kinase buffer. The  $^{32}\text{P}$  incorporated into the substrate is separated from free  $^{32}\text{P}$ -ATP by electrophoresis and the incorporated  $^{32}\text{P}$  is counted using a radioisotope counter. The amount of incorporated  $^{32}\text{P}$  is proportional to the activity of XMAD. A determination of

WO 01/42285

PCT/US00/32990

linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

5 A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

#### IX. Complementary Polynucleotides

10 Sequences complementary to the XMAD-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring XMAD. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of XMAD. To inhibit transcription, a  
15 complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the XMAD-encoding transcript.

#### X. Expression of XMAD

Expression and purification of XMAD is achieved using bacterial or virus-based expression  
20 systems. For expression of XMAD in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3).  
25 Antibiotic resistant bacteria express XMAD upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of XMAD in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding XMAD by either homologous recombination or bacterial-mediated  
30 transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther.  
35 7:1937-1945.)



WO 01/42285

PCT/US00/32990

the specific amino acid residue phosphorylated is made by phosphoamino acid analysis of the hydrolyzed protein.

## XII. Functional Assays

XMAD function is assessed by expressing the sequences encoding XMAD at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT plasmid (Life Technologies) and pCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of XMAD on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding XMAD and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding XMAD and other genes of interest can be analyzed by northern analysis or microarray techniques.

WO 01/42285

PCT/US00/32990

### XIII. Production of XMAD Specific Antibodies

XMAD substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

5 Alternatively, the XMAD amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, *supra*, ch. 11.)

10 Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for anti-peptide and anti-XMAD  
15 activity by, for example, binding the peptide or XMAD to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

### XIV. Purification of Naturally Occurring XMAD Using Specific Antibodies

Naturally occurring or recombinant XMAD is substantially purified by immunoaffinity chromatography using antibodies specific for XMAD. An immunoaffinity column is constructed by  
20 covalently coupling anti-XMAD antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing XMAD are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of XMAD (e.g., high ionic strength  
25 buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/XMAD binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and XMAD is collected.

### XV. Identification of Molecules Which Interact with XMAD

XMAD, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent.  
30 (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled XMAD, washed, and any wells with labeled XMAD complex are assayed. Data obtained using different concentrations of XMAD are used to calculate values for the number, affinity, and association of XMAD with the candidate molecules.

WO 01/42285

PCT/US00/32990

Alternatively, molecules interacting with XMAD are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989, Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

XMAD may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention.

Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

WO 01/42285

PCT/US00/32990

Table I

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	22	1424691	BEPINON01	998379R1 (KIDNTUT01), 1424691H1 (BEPINON01), SXAEO2538V1
2	23	1450801	PENITUT01	046316H1 (CORNNOT01), 1450801CT1 (PENITUT01), 1450801H1 (PENITUT01), 1671961H1 (BLADNOT05)
3	24	1597872	BRAINOT14	814997R1 (OVARUTUT01), 814997T1 (OVARUTUT01), 1412857T6 (BRAINOT12), 1438406F1 (PANCNOT08), 1597872H1 (BRAINOT14), 1797683H1 (PROSTUT05), 3346671H1 (BRAITUT24)
4	25	1674661	BLADNOT05	1655227F6 (PROSTUT08), 1674661H1 (BLADNOT05), 1675239F6 (BLADNOT05), 1879940F6 (LEUKNOT03), 2121172F6 (BRSTNOT07), 2157008F6 (BRAINOT09), 2672389F6 (KIDNNOT19), 3270393H1 (BRAINOT20), 3387668H1 (LUNGUTUT17), 3685486H1 (HEAANOT01), 4103531H1 (BRSTTUT17), 4850546H1 (TESTNOT10), 5028429H1 (COLCDIT01), 5661414H1 (BRAUNOT01)
5	26	1689337	PROSTUT10	542204R1 (OVARNOT02), 961047R2 (BRSTTUT03), 1623395T6 (BRAITUT13), 1663607F6 (BRSTNOT09), 1689337H1 (PROSTUT10), 2898863R6 (THYMNON04), 3507526H1 (CONCNOT01)
6	27	1746392	STOMTUT02	682990H1 (UTRSNOT02), 1663009F6 (BRSTNOT09), 1746392H1 (STOMTUT02), 1746392T6 (STOMTUT02), 2079257F6 (ISLTNOT01), 3099537F6 (PTHYNOT03), 3111943H1 (BRSTNOT17), 3391682H1 (LUNGNOT28), 474243F6 (SMCRUNT01)
7	28	1825182	LSUBNOT03	983441H1 (TONGTUT01), 1825182F6 (LSUBNOT03), 1825182H1 (LSUBNOT03), 1825369F6 (LSUBNOT03), SAQB00477F1, SAQB00879F1, SAQB01310F1, SAQB00187F1, SAQA02142F1, SAQA00159F1
8	29	2155541	BRAINOT09	871127T1 (LUNGAST01), 1309342R1 (COLNFET02), 1544021T1 (PROSTUT04), 2155541H1 (BRAINOT09), 2155541X15F1 (BRAINOT09)

WO 01/42285

PCT/US00/32990

Table 1 (Cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
9	30	2215706	SINTFET03	570718H1 (MMLR3DT01), 756160R1 (BRAITUT02), 1511501F1 (LUNGNOT14), 2215706F6 (SINTFET03), 2215706H1 (SINTFET03), 2648753F6 (OVARNOT10), 2804428H1 (PENCNOT01), 3092672T6 (BRSTNOT19), 3597972H1 (FIBPNOT01), 3604953H1 (LUNGNOT30), 3879505H1 (SPLNNOT11), 4506625F6 (OVARDT01), 4708188H1 (BRAIFET02), 4985942H1 (LIVRTUT10), 5151981H1 (HEARFET03), 5644791H1 (UTRSTMR01), 5862219H1 (BRAYDIT01)
10	31	2347692	TESTTUT02	075856R1 (THPIPEB01), 370791R1 (LUNGNOT02), 1502478F1 (BRAITUT07), 2347692H1 (TESTTUT02), 2825041F6 (ADRETUT06)
11	32	2579048	KIDNTUT13	841019R1 (PROSTUT05), 1352253F1 (LATRTUT02), 1414589F6 (BRAINOT12), 1427648F1 (SINTBST01), 1996374R6 (BRSTTUT03), 2579048H1 (KIDNTUT13)
12	33	2604493	LUNGTUT07	901679X18 (BRSTTUT03), 927970X54R1 (BRAINOT04), 1435427F6 (PANCNOT08), 1484806F6 (CORPNOT02), 1962694T6 (BRSTNOT04), 1990921F6 (CORPNOT02), 2279985R6 (PROSNON01), 2279985T6 (PROSNON01), 2294223T6 (BRAINON01), 2604493H1 (LUNGTUT07), 2707717H1 (PONSATZ01), 3421936H1 (UCMCNOT04), 4769752H1 (BRAFNOT02), 4989101H1 (LIVRTUT10), SAEA01968R1
13	34	2787182	BRSTNOT13	916228H1 (BRSTNOT04), 1251624F6 (LUNGFET03), 1440454F6 (THYRNOT03), 1664091F6 (BRSTNOT09), 1812788X17C1 (PROSTUT12), 1812788X21C1 (PROSTUT12), 2787182H1 (BRSTNOT13), 2824324F6 (ADRETUT06), 2882479T6 (UTRSTUT05), 3833543H1 (PANCNOT17), 60147041D2, 60147042B6, 60147044D2, SXAE05916V1, SXAE02927V1, g1670173
14	35	3096668	CERVNOT03	2373962F6 (ISLTNOT01), 2373962T6 (ISLTNOT01), 2762988H1 (BRSTNOT12), 3096668F6 (CERVNOT03), 3096668H1 (CERVNOT03), 3096668T6 (CERVNOT03), SCGA06156V1, SCGA11275V1, SCGA07741V1

WO 01/42285

PCT/US00/32990

Table 1 (Cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
15	36	3143411	HNT2AZS07	540532T6 (LNODNOT02), 852710R1 (NGANNOT01), 860567R1 (BRAITUT03), 1402106F6 (LATRTUT02), 3143411H1 (HNT2AZS07), 3143411R6 (HNT2AZS07), 5135819H1 (OVARADIT04)
16	37	3170835	BRSTNOT18	3170835H1 (BRSTNOT18), 3171275F6 (BRSTNOT18)
17	38	3550808	SYNONOT01	00101F1 (U937NOT01), 1353706T1 (LATRTUT02), 1426227F1 (SINTBST01), 1804230F6 (SINTNOT13), 2361183T6 (LUNGFET05), 2606392H1 (LUNGUTUT07), 3550808H1 (SYNONOT01), SBAA00101F1
18	39	3683905	HEAANOT01	833556H1 (PROSNOT07), 1494051H1 (PROSNON01), 3683905H1 (HEAANOT01), 5512558H1 (BRADDIR01), 5700822H1 (DRGCNOT01), 91267581
19	40	4062841	BRAINOT21	1863239H1 (PROSNOT19), 1863239T6 (PROSNOT19), 4062841H1 (BRAINOT21)
20	41	6394358	UTRENOT10	875733R6 (LUNGAST01), 1312637T6 (BLADTUT02), 2296386R6 (BRSTNOT05), 2296386T6 (BRSTNOT05), 6394358H1 (UTRENOT10)
21	42	2847752		94126329.v113.gs_6.edit.5p 1-6416; g3449297 5802-10044 5547763H1 (TESTNOC01) 8322-8532; 3373379H1 (CONNTUT05) 8722-8984; 3331371H1 (BRAIFET01) 9378-9641; 5376974H1 (BRAXNOT01) 9634-9772; 4015537F6 (BRAXNOT01) 9709-10226; 5921447H1 (BRAIFET02) 9963-10235; 4700084F6 (BRALNOT01) 10218-10739; 670937H1 (CRBLNOT01) 10517-10780; 3788576H1 (BRAHNOT05) 10608-10904; 5929111H1 (BRAIFET02) 10791-11069; 3278762T6 (STOMFET02) 11029-11627; 2847752R6 (HNT2AZS07) 11181-11648

WO 01/42285

PCT/US00/32990

Table 2

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Homologous Sequence	Analytical Methods and Databases
1 (1424691)	222	S25 T88 S155 S179 T191 T207 S90 S96 S113 S119	N7 N53 N68	M1-G29: Signal peptide E173-E181: 5- hydroxytryptamine 5A (serotonin) receptor R213-D215: Rgd cell interaction motif	g7649266 Sterile-alpha motif and leucine zipper containing kinase AZK	MOTIFS BLAST_GENBANK SPSCAN BLIMPS-PRINTS
2 (1450801)	228	S177 S99	N25	F20-G211: Leucine Rich Repeat	g3786312 Extracellular matrix protein (Nishiu, J. et al. Genomics (1998) 52:378- 381)	MOTIFS BLAST_GENBANK HMMER PFAM BLIMPS-PRINTS
3 (1597872)	386	S88 S137 S229 S364 T370		G237-P247: Insulin- like growth factor G202-R341: Spliceosome- associated protein	g4033606 Extensin (Kieliszewski M.J., Lamport D.T. Plant J. (1994) 5:157- 172)	MOTIFS BLAST_GENBANK BLIMPS-BLOCKS BLAST-DOMO

WO 01/42285

PCT/US00/32990

Table 2 (Cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Homologous Sequence	Analytical Methods and Databases
4 (1674661)	833	S46 T276 T418 S34 T57 T229 T302 S382 T429 S505 S826 S200 S364 S480 T523 S555 T561 S685 T701 S742 Y249 Y345 Y736	N106 N121 N310 N419 N522 N564	M1-V23: Signal peptide F53-K481: Semaphorin domain	g1110599 Semaphorin homolog (Inagaki, S. et al. FEBS Lett (1995) 370:269-272)	MOTIFS BLAST_GENBANK SIGPEPT SPSCAN BLAST-PFAM
5 (1689337)	410	T358 S394 S139 S249 S17 S343 T385 S401	N3	M1-A16: Signal peptide P240-Q255: Prokaryotic molybdopterin oxidoreductase L55-A65: Prepro- orexin signature R349-D351: Rgd cell interaction motif	g3450883 Fibroin (Gosline, J.M. et al. J. Exp. Biol. (1999) 202:3295-3303)	MOTIFS BLAST_GENBANK SPSCAN BLIMPS-BLOCKS BLIMPS-PRINTS
6 (1746392)	360	S217 S255 T344 S37 S8 T28 S69 S113 T182 S188 S224 S242 S250		E17-P356: Neurofilament triplet H	g310200 proline-rich proteoglycan	MOTIFS BLAST_GENBANK BLAST-DOMO
7 (1825182)	377	S18 S41 S140 T267 S38 S62 S120 T343	N97 N128 N135 N146	F20-G211: Leucine Rich Repeat	g188864 mucin (Shimomura, T., Blood (1990) 75:2349-2356)	MOTIFS BLAST_GENBANK HMMER-PFAM BLIMPS-PRINTS



WO 01/42285

PCT/US00/32990

Table 2 (Cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Homologous Sequence	Analytical Methods and Databases
8 (2155541)	182	T116		M1-F22: Signal peptide T42-D62: Neutrophil cytosol factor D170-E178: 7-fold repeat proteins I motif R60-D62: Rgd cell interaction motif	g6164953 vacuolar sorting protein VPS29 [Mus musculus] Edgar, A.J. and Polak, J.M. (2000) Biochem. Biophys. Res. Commun. 277:622-630	MOTIFS BLAST_GENBANK SPSCAN BLIMPS-PRINTS BLIMPS-PFAM
9 (2215706)	513	S239 S325 T14 S51 S71 S110 S137 S171 S208 S213 T219 S263 S268 S349 S394 S403 S404 T96 T118 S149 S239 S365 T475 S479	N69 N89 N338 N429	S94-Q108: Adrenocorticotropi n receptor S328-S335: "Phage" integrase family R196-D198: Rgd cell interaction motif	g4322670 Dentin phosphoryn P = 7.6e-07	MOTIFS BLIMPS-PRINTS BLIMPS-PFAM
10 (2347692)	361	T81 T53 S158 S257 T333 S128	N8 N51 N306 N324	G242-M268: C1q domain proteins L146-P150: Laminin G domain protein G110-P288: Fibrillar collagen carboxyl-terminus	g1562534 csdp single- stranded DNA binding protein	MOTIFS BLAST_GENBANK BLIMPS-BLOCKS BLIMPS-PFAM BLAST-DOMO
11 (2579048)	327	S18 S199 T55 T72 S73 T285 S51 S140 S177 S262	N158	K41-T55: Histone H5 signature R213-D215: Rgd cell interaction motif		MOTIFS BLIMPS-PRINTS

WO 01/42285

PCT/US00/32990

Table 2 (Cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Homologous Sequence	Analytical Methods and Databases
12 (2604493)	1 110	S213 T795 S8 T17 S40 S64 S82 T118 S132 S151 S170 S192 S249 T317 T357 T476 S601 S642 S673 S674 S701 S731 T795 S1015 S1059 S1073 T84 S236 S276 T292 T309 S337 S457 T506 T744 S749 S813 T945 S976 T1078	N168 N472 N640 N671 N672 N691 N698 N729 N747 N851 N966	V532-V565, L577- F598, M609-L622: Armadillo/beta- catenin-like repeats S511-H1025: Mouse p120 protein R971-D973: Rgd cell interaction motif	g1702924 p0071 Catenin- related protein (Hatzfeld M., Nachtsheim C. J. Cell Sci. (1996) 109:2767-2778)	MOTIFS BLAST_GENBANK HMMER-PFAM BLIMPS-PFAM BLAST-PRODOM
13 (2787182)	386	S123 T210 S265 S283 S317 S326 T330 T338 T44 S79 S100 S109 T127 T142 T170 T214 S332	N14 N173	C57-Q72: Transmembrane motif M279-M363: osteonectin	g3876060 Weak similarity with nitrogen fixation regulator	MOTIFS BLAST_GENBANK HMMER PROFILESSCAN
14 (3096668)	181	S114 S116 S118 S120 S122 T124 S154 T173 T13	N17	F16-E25: Alpha-type calcitonin signature	g3393011 Clumping factor B (Ni Eidhin D. et al. Mol. Microbiol. 1998:30:245- 257)	MOTIFS BLAST_GENBANK BLIMPS-PRINTS

WO 01/42285

PCT/US00/32990

Table 2 (Cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Homologous Sequence	Analytical Methods and Databases
15 (3143411)	374	S291 S75 S92 T206 T214 T298 T315 T23 T37 T50 S51 S262 T263 S306	N109 N304	M1-G24: Signal peptide N224-W247, V328- N348: Transmembrane motif Y46-C63, W163-C176: C-type lectin domain	g3790610 Layilin (Borowsky M.L., Hynes R.O. (1998) J. Cell Biol. 143:429-442)	MOTIFS BLAST_GENBANK SIGPEPT SPSCAN HMMER BLAST PFAM BLIMPS-PRINTS PROFILES-SCAN BLAST-DOMO
16 (3170835)	102	S45 T57 T44	N33	M1-T19: Signal peptide A55-103: Insect cuticle protein	g2565394 Cuticle 12	MOTIFS BLAST_GENBANK SPSCAN BLAST-PFAM PROFILES-SCAN BLIMPS-PRINTS BLAST-PRODOME BLAST-DOMO
17 (3550808)	510	S76 S88 S128 S150 T152 T308 T448 T461 S49 S56 S110 T138	N72 N136 N193 N253 N352 N411	M1-G20: Signal peptide L5-G23: Transmembrane motif G33-F46: Pheromone B alpha-1 receptor	g294502 Olfactomedin (Yokoe H., Anholt R.R. Proc. Natl. Acad. Sci. USA (1993) 90:4655-9)	MOTIFS BLAST_GENBANK SIGPEPT SPSCAN HMMER BLIMPS-PRINTS

WO 01/42285

PCT/US00/32990

Table 2 (Cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Homologous Sequence	Analytical Methods and Databases
18 (3683905)	185	S11 S57 S173 T135		M1-L170: von Willebrand factor domain score M2-F15, R37-F51, V103-G111 M1-R171: collagen glycoprotein precursor	g2654431 Type XII collagen	MOTIFS BLAST_GENBANK HMMER-PFAM BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO
19 (4062841)	207	T83 S201 T74 S166 T174 S190		V93-T174: PDZ domain (Also known as DHR or GLGF) P91-V171: SH3 domain	g3885828 Lin-7-A (Irie M. et al. Oncogene (1999) 18:2811-2817)	MOTIFS BLAST_GENBANK HMMER-PFAM BLIMPS-PRINTS BLIMPS-PFAM BLIMPS-PRODOM BLAST-PRODOM BLAST-DOMO
20 (6394358)	238	S2 S96 S100 S12 S26 T149 S200 T203	N160	L3-A25, T64-E110, E209-R218: 7-fold repeat proteins (clathrin) R136-D138: Rgd cell interaction motif		MOTIFS BLIMPS-PFAM

WO 01/42285

PCT/US00/32990

Table 2 (Cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Homologous Sequence	Analytical Methods and Databases
21	3298	S93 T307 S2706 S29 S222 S240 T266 T403 T421 T440 T457 T533 T567 T569 S658 T744 T756 T778 T831 S865 S880 S882 T931 T944 S983 T985 S1030 T1089 T1223 T1246 T1304 S1347 S1366 T1433 T1465 T1488 S1578 S1630 S1685 T1841 S1903 T2184 S2359 S2368 T2482 S2513 S2663 S2847 T2863 S2878 S2880 S2885 S2897 T2964 T2979 S2980 S3053 S3281 S186 S194 T200 T250 S409 S497 T527 S545 S833 T852 T1020 S1180 S1230 T1275 S1659 S1660 S1737 S1829 S1976 T2066 S2113 T2216 S2322 S2355 S2549 S2552 T2709 S2819 S2822 S2830 S2919 S2956 T2964 T3273 Y399 Y449 Y842 Y2299	N631 N846 N1181 N1221 N1316 N1326 N1648 N1712 N1769 N2034 N2163 N2182 N2372 N2460 N2492 N2683 N2732 N2794 N3240	L1536-V1553 L2525-A2545 I2669-L2689 S2712-A2730 L2741-N2761: Transmembrane Domains M1-E31: Signal peptide Y329-A423; Y437- L535 Y549-V641; F655- L746 Y760-T848; Y862- N952 Y965-Q1057; F1071- V1159 Y1178-I1265: Cadherin domains C1378-C1431; C1438- C1469 C1478-C1512; C1725- C1756 C1931-C1962; C1966- C2000: EGF domains F1542-D1704; F1792- E1920: Laminin G domains C2110-A2137; C2588- L2613: GPCR signature	g3449288 MEGF2 [Rattus norvegicus]	MOTIFS BLAST_GENBANK HMMER SPSCAN HMMER-PFAM BLIMPS-BLOCKS PROFILESAN BLIMPS-PRINTS

WO 01/42285

PCT/US00/32990

Table 3

Polynucleotide SEQ ID NO:	Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
22	1-517 852-905	Cardiovascular (0.200) Gastrointestinal (0.200) Urologic (0.222)	Cell proliferation(0.400) Cell proliferation/Cell line (0.200) Inflammation/Trauma(0.300)	pSPORT1
23	1-2387	Reproductive (0.397) Gastrointestinal (0.132) Musculoskeletal (0.118)	Cell proliferation(0.559) Inflammation/Trauma(0.176) Other (0.118)	pINCY
24	1-901 1456-1471	Reproductive (0.262) Gastrointestinal (0.179) Nervous (0.179)	Cell proliferation(0.536) Inflammation/Trauma(0.297) Cell proliferation/Cell line (0.190)	pINCY
25	1-1928 1776-3293	Reproductive (0.294) Gastrointestinal (0.157) Nervous (0.157)	Cell proliferation(0.471) Inflammation/Trauma(0.373) Cell proliferation/Cell line (0.118)	pINCY
26	1-821 1312-1324	Reproductive (0.351) Hematopoietic/Immune (0.135) Nervous (0.135)	Cell proliferation(0.486) Cell proliferation/Cell line (0.243) Inflammation/Trauma(0.351)	pINCY
27	1-626 1034-1324	Reproductive (0.227) Nervous (0.182) Gastrointestinal (0.159)	Cell proliferation(0.432) Cell proliferation/Cell line (0.205) Inflammation/Trauma(0.228)	pINCY
28	1-2429	Gastrointestinal (1.000)	Cell proliferation(1.000)	pINCY
29	1-50 591-985	Reproductive (0.197) Nervous (0.164) Hematopoietic/Immune (0.145)	Cell proliferation(0.441) Inflammation/Trauma(0.454) Cell proliferation/Cell line (0.191)	pINCY
30	1-285 813-930 1145-3381	Nervous (0.235) Reproductive (0.235) Gastrointestinal (0.126)	Cell proliferation(0.445) Cell proliferation/Cell line (0.202) Inflammation/Trauma(0.252)	pINCY
31	1-82 1728-1803	Reproductive (0.267) Nervous (0.233) Other (0.117)	Cell proliferation(0.383) Inflammation/Trauma(0.383) Cell proliferation/Cell line (0.217)	pINCY
32	1-430 752-964 1405-1515	Reproductive (0.263) Nervous (0.193) Cardiovascular (0.140)	Cell proliferation(0.544) Cell proliferation/Cell line (0.158) Inflammation/Trauma(0.246)	pINCY

WO 01/42285

PCT/US00/32990

Table 3 (Cont.)

Polynucleotide SEQ ID NO:	Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
33	1-570 860-1573 1859-2494 2864-4416	Nervous (0.383) Reproductive (0.210) Gastrointestinal (0.111)	Cell proliferation(0.395) Inflammation/Trauma(0.334) Cell proliferation/Cell line (0.198)	pINCY
34	1-30 190-234 889-4428	Reproductive (0.280) Nervous (0.195) Hematopoietic/Immune (0.146)	Cell proliferation(0.415) Inflammation/Trauma(0.366) Cell proliferation/Cell line (0.146)	pINCY
35	1-189 264-1907	Reproductive (0.333) Cardiovascular (0.167) Developmental (0.167)	Cell proliferation(0.333) Cell proliferation/Cell line (0.333) Inflammation/Trauma(0.375)	pINCY
36	1-773 1742-1839	Nervous (0.312) Reproductive (0.312) Gastrointestinal (0.125)	Cell proliferation(0.531) Inflammation/Trauma(0.250) Cell proliferation/Cell line (0.125)	pSPORT1
37	1-503	Reproductive (1.000)	Cell proliferation(1.000)	pINCY
38	1-167 449-946 1541-2154	Gastrointestinal (0.723) Reproductive (0.128) Urologic (0.085)	Cell proliferation(0.447) Inflammation/Trauma(0.489) Trauma (0.170)	pINCY
39	1-431 666-733	Reproductive (0.450) Nervous (0.250) Urologic (0.100)	Cell proliferation(0.750) Inflammation/Trauma(0.200) Trauma (0.100)	pINCY
40	1-48 301-453 634-665	Reproductive (0.300) Cardiovascular (0.200) Hematopoietic/Immune (0.200)	Cell proliferation(0.400) Cell proliferation/Cell line (0.300) Inflammation/Trauma(0.400)	pINCY
41	1-276 553-741 820-1235	Hematopoietic/Immune (0.349) Nervous (0.163) Gastrointestinal (0.140)	Inflammation/Trauma(0.512) Cell proliferation(0.349) Cell proliferation/Cell line (0.209)	pINCY

WO 01/42285

PCT/US00/32990

Table 4

Polynucleotide SEQ ID NO:	Library	Library Comment
22	BEPINON01	This normalized bronchial epithelium library was constructed from 5.12 million independent clones from the BEPINOT01 library. RNA was made from a bronchial epithelium primary cell line derived from a 54-year-old Caucasian male. The normalization and hybridization conditions were adapted from Soares et al., PNAS (1994) 91:9228, using a longer (24-hour) reannealing hybridization period.
23	PENITUT01	Library was constructed using RNA isolated from tumor tissue removed from the penis of a 64-year-old Caucasian male during penile amputation. Pathology indicated a fungating invasive grade 4 squamous cell carcinoma involving the inner wall of the foreskin and extending onto the glans penis. Patient history included benign neoplasm of the large bowel, atherosclerotic coronary artery disease, angina pectoris, gout, and obesity. Family history included malignant pharyngeal neoplasm, chronic lymphocytic leukemia, and chronic liver disease.
24	BRAINOT14	Library was constructed using RNA isolated from brain tissue removed from the left frontal lobe of a 40-year-old Caucasian female during excision of a cerebral meningial lesion. Pathology for the associated tumor tissue indicated grade 4 gemistocytic astrocytoma.
25	BLADNOT05	Library was constructed using RNA isolated from bladder tissue removed from a 60-year-old Caucasian male during a radical cystectomy, prostatectomy, and vasectomy. Pathology for the associated tumor tissue indicated grade 3 transitional cell carcinoma. Carcinoma in-situ was identified in the dome and trigone. Patient history included tobacco use.
26	PROSTUT10	Library was constructed using RNA isolated from prostatic tumor tissue removed from a 66-year-old Caucasian male during radical prostatectomy and regional lymph node excision. Pathology indicated an adenocarcinoma (Gleason grade 2+3). Adenofibromatous hyperplasia was also present. The patient presented with elevated prostate specific antigen (PSA). Family history included prostate cancer and secondary bone cancer.
27	STOMTUT02	Library was constructed using RNA isolated from stomach tumor tissue obtained from a 68-year-old Caucasian female during a partial gastrectomy. Pathology indicated a malignant lymphoma of diffuse large-cell type. Previous surgeries included cholecystectomy. Patient history included thalassemia. Family history included acute leukemia, malignant neoplasm of the esophagus, malignant stomach neoplasm, and atherosclerotic coronary artery disease.



WO 01/42285

PCT/US00/32990

Table 4 (Cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
28	LSUBNOT03	Library was constructed using RNA isolated from submandibular gland tissue obtained from a 68-year-old Caucasian male during a sialoadenectomy. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
29	BRAINOT09	Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus, who died at 23 weeks' gestation.
30	SINTFET03	Library was constructed using RNA isolated from small intestine tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
31	TESTTUT02	Library was constructed using RNA isolated from testicular tumor removed from a 31-year-old Caucasian male during unilateral orchiectomy. Pathology indicated embryonal carcinoma.
32	KIDNTUT13	Library was constructed using RNA isolated from kidney tumor tissue removed from a 51-year-old Caucasian female during a nephroureterectomy. Pathology indicated a grade 3 renal cell carcinoma. Patient history included depressive disorder, hypoglycemia, and uterine endometriosis. Family history included calculus of the kidney, colon cancer, and type II diabetes.
33	LUNGUT07	Library was constructed using RNA isolated from lung tumor tissue removed from the upper lobe of a 50-year-old Caucasian male during segmental lung resection. Pathology indicated an invasive grade 4 squamous cell adenocarcinoma. Patient history included tobacco use. Family history included skin cancer.
34	BRSTNOT13	Library was constructed using RNA isolated from breast tissue removed from the left medial lateral breast of a 36-year-old Caucasian female during bilateral simple mastectomy and total breast reconstruction. Pathology indicated benign breast tissue. Patient history included a breast neoplasm, depressive disorder, hyperlipidemia, chronic stomach ulcer, and an ectopic pregnancy. Family history included myocardial infarction, cerebrovascular disease, atherosclerotic coronary artery disease, hyperlipidemia, skin cancer, breast cancer, depressive disorder, esophageal cancer, bone cancer, Hodgkin's lymphoma, bladder cancer, and heart condition.
35	CERVNOT03	Library was constructed using RNA isolated from uterine cervical tissue removed from a 40-year-old Caucasian female during a vaginal hysterectomy with bilateral salpingo-oophorectomy and dilation and curettage. Pathology indicated secretory phase endometrium.

WO 01/42285

PCT/US00/32990

Table 4 (Cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
36	HNT2AZS07	This subtracted library was constructed from RNA isolated from an hNT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor) treated for three days with 0.35 micromolar AZ. The hybridization probe for subtraction was derived from a similarly constructed library from untreated hNT2 cells. 3.08M clones from the AZ-treated library were subjected to three rounds of subtractive hybridization with 3.04M clones from the untreated library. Subtractive hybridization conditions were based on the methodologies of Swaroop et al. (NAR (1991) 19:1954) and Bonaldo et al. (Genome Research (1996) 6:791).
37	BRSTNOT18	Library was constructed using RNA isolated from diseased breast tissue removed from a 57-year-old Caucasian female during a unilateral simple extended mastectomy. Pathology indicated mildly proliferative breast disease. Patient history included breast cancer and osteoarthritis. Family history included type II diabetes, gallbladder and breast cancer, and chronic lymphocytic leukemia.
38	SYNONOT01	Library was constructed using RNA isolated from synovial tissue removed from a 75-year-old Caucasian male.
39	HEANOT01	Library was constructed using RNA isolated from right coronary and right circumflex coronary artery tissue removed from the explanted heart of a 46-year-old Caucasian male during a heart transplantation. Patient history included myocardial infarction from total occlusion of the left anterior descending myocardial atherosclerotic coronary artery disease, hyperlipidemia, myocardial ischemia, dilated cardiomyopathy, left ventricular dysfunction, and tobacco abuse. Previous surgeries included cardiac catheterization. Family history included atherosclerotic coronary artery disease.
40	BRAINOT21	Library was constructed using RNA isolated from diseased brain tissue removed from the left frontal lobe of a 46-year-old Caucasian male during a lobectomy. Pathology indicated focal cortical and subcortical scarring of the left frontal lobe, characterized by cavitation and extensive reactive changes, including marked gliosis and hemosiderin deposition, consistent with a history of remote severe head trauma. GFAP was positive in astrocytes. The pattern of reactivity is that of reactive gliosis. Patient history included traumatic intracranial hemorrhage and brain injury with loss of consciousness following head trauma. Family history included cerebrovascular disease, cerebrovascular disease, and atherosclerotic coronary artery disease.

WO 01/42285

PCT/US00/32990

Table 4 (Cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
41	UTRENOT10	Library was constructed using polyA RNA isolated from pooled uterine endometrial tissue removed from three adult females during endometrial biopsy. Pathology indicated normal endometrium.
42		The Incyte cDNAs for SEQ ID NO:42 were derived from cDNA libraries constructed from brain, including tissues associated with Huntington's disease, Alzheimer's disease, and multiple sclerosis, as well as from pituitary, testicular, stomach, spinal cord, kidney, and prostate tissues, and from ovarian, cervical, pancreatic, and soft tissue tumors.

WO 01/42285

PCT/US00/32990

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%.
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families

WO 01/42285

PCT/US00/32990

Table 5 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Normalized quality score $\geq$ GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. <u>supra</u> ; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

WO 01/42285

PCT/US00/32990

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

- 5       a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21,  
      b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21,  
      c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and  
10       d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-21.

15

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

20

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:22-42.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

25

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

30

9. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

35

b) recovering the polypeptide so expressed.

WO 01/42285

PCT/US00/32990

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

- 5 a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42,
- b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42,
- c) a polynucleotide sequence complementary to a),
- d) a polynucleotide sequence complementary to b), and
- 10 e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization
- 20 complex is formed between said probe and said target polynucleotide or fragments thereof, and
- b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

25

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- 30 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

35

WO 01/42285

PCT/US00/32990

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

18. A method for treating a disease or condition associated with decreased expression of functional XMAD, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional XMAD, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional XMAD, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a



WO 01/42285

PCT/US00/32990

Leu	Asn	Tyr	Asn	Pro	Phe	Asp	Gln	Lys	Leu	Tyr	Val	Tyr	Asn	Asp
				485					490					495
Gly	Tyr	Leu	Leu	Asn	Tyr	Asp	Leu	Ser	Val	Leu	Gln	Lys	Pro	Gln
				500					505					510

<210> 18  
 <211> 185  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3683905CD1

*fragment of human Wnt1 (about 1/2 of protein)*

<400> 18	Met	Phe	Leu	Leu	Asp	Ser	Ser	Ala	Ser	Val	Ser	His	Tyr	Glu	Phe
	1				5					10					15
	Ser	Arg	Val	Arg	Glu	Phe	Val	Gly	Gln	Leu	Val	Ala	Pro	Leu	Pro
					20					25					30
	Leu	Gly	Thr	Gly	Ala	Leu	Arg	Ala	Ser	Leu	Val	His	Val	Gly	Ser
					35					40					45
	Arg	Pro	Tyr	Thr	Glu	Phe	Pro	Phe	Gly	Gln	His	Ser	Ser	Gly	Glu
					50					55					60
	Ala	Ala	Gln	Asp	Ala	Val	Arg	Ala	Ser	Ala	Gln	Arg	Met	Gly	Asp
					65					70					75
	Thr	His	Thr	Gly	Leu	Ala	Leu	Val	Tyr	Ala	Lys	Glu	Gln	Leu	Phe
					80					85					90
	Ala	Glu	Ala	Ser	Gly	Ala	Arg	Pro	Gly	Val	Pro	Lys	Val	Leu	Val
					95					100					105
	Trp	Val	Thr	Asp	Gly	Gly	Ser	Ser	Asp	Pro	Val	Gly	Pro	Pro	Met
					110					115					120
	Gln	Glu	Leu	Lys	Asp	Leu	Gly	Val	Thr	Val	Phe	Ile	Val	Ser	Thr
					125					130					135
	Gly	Arg	Gly	Asn	Phe	Leu	Glu	Leu	Ser	Ala	Ala	Ala	Ser	Ala	Pro
					140					145					150
	Ala	Glu	Lys	His	Leu	His	Phe	Val	Asp	Val	Asp	Asp	Leu	His	Ile
					155					160					165
	Ile	Val	Gln	Glu	Leu	Arg	Gly	Ser	Ile	Leu	Asp	Ala	Met	Arg	Pro
					170					175					180
	Gln	Ala	Tyr	Ser	Leu										
					185										

<210> 19  
 <211> 207  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 4062841CD1

<400> 19	Met	Ala	Ala	Leu	Val	Glu	Pro	Leu	Gly	Leu	Glu	Arg	Asp	Val	Ser
	1				5					10					15
	Arg	Ala	Val	Glu	Leu	Leu	Glu	Arg	Leu	Gln	Arg	Ser	Gly	Glu	Leu
					20					25					30
	Pro	Pro	Gln	Lys	Leu	Gln	Ala	Leu	Gln	Arg	Val	Leu	Gln	Ser	Arg
					35					40					45
	Phe	Cys	Ser	Ala	Ile	Arg	Glu	Val	Tyr	Glu	Gln	Leu	Tyr	Asp	Thr
					50					55					60
	Leu	Asp	Ile	Thr	Gly	Ser	Ala	Glu	Ile	Arg	Ala	His	Ala	Thr	Ala
					65					70					75
	Lys	Ala	Thr	Val	Ala	Ala	Phe	Thr	Ala	Ser	Glu	Gly	His	Ala	His
					80					85					90
	Pro	Arg	Val	Val	Glu	Leu	Pro	Lys	Thr	Asp	Glu	Gly	Leu	Gly	Phe
					95					100					105

WO 01/42285

PCT/US00/32990

compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

- 5 a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- 10 c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- 15 a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- 20 c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

- a) treating a biological sample containing nucleic acids with the test compound;
- 25 b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;
- 30 c) quantifying the amount of hybridization complex; and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

Asn	Ile	Met	Gly	Gly	Lys	Glu	Gln	Asn	Ser	Pro	Ile	Tyr	Ile	Ser	
				110					115					120	
Arg	Val	Ile	Pro	Gly	Gly	Val	Ala	Asp	Arg	His	Gly	Gly	Leu	Lys	
				125					130					135	
Arg	Gly	Asp	Gln	Leu	Leu	Ser	Val	Asn	Gly	Val	Ser	Val	Glu	Gly	
				140					145					150	
Glu	Gln	His	Glu	Lys	Ala	Val	Glu	Leu	Leu	Lys	Ala	Ala	Gln	Gly	
				155					160					165	
Ser	Val	Lys	Leu	Val	Val	Arg	Tyr	Thr	Pro	Arg	Val	Leu	Glu	Glu	
				170					175					180	
Met	Glu	Ala	Arg	Phe	Glu	Lys	Met	Arg	Ser	Ala	Arg	Arg	Arg	Gln	
				185					190					195	
Gln	His	Gln	Ser	Tyr	Ser	Ser	Leu	Glu	Ser	Arg	Gly				
				200					205						

&lt;210&gt; 20

&lt;211&gt; 238

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6394358CD1

&lt;400&gt; 20

Met	Ser	Leu	Asn	Glu	His	Ser	Met	Gln	Ala	Leu	Ser	Trp	Arg	Lys	
1				5					10					15	
Leu	Tyr	Leu	Ser	Arg	Ala	Lys	Leu	Lys	Ala	Ser	Ser	Arg	Thr	Ser	
				20					25					30	
Ala	Leu	Leu	Ser	Gly	Phe	Ala	Met	Val	Ala	Met	Val	Glu	Val	Gln	
				35					40					45	
Leu	Asp	Ala	Asp	His	Asp	Tyr	Pro	Pro	Gly	Leu	Leu	Ile	Ala	Phe	
				50					55					60	
Ser	Ala	Cys	Thr	Thr	Val	Leu	Val	Ala	Val	His	Leu	Phe	Ala	Leu	
				65					70					75	
Met	Ile	Ser	Thr	Cys	Ile	Leu	Pro	Asn	Ile	Glu	Ala	Val	Ser	Asn	
				80					85					90	
Val	His	Asn	Leu	Asn	Ser	Val	Lys	Glu	Ser	Pro	His	Glu	Arg	Met	
				95					100					105	
His	Arg	His	Ile	Glu	Leu	Ala	Trp	Ala	Phe	Ser	Thr	Val	Ile	Gly	
				110					115					120	
Thr	Leu	Leu	Phe	Leu	Ala	Glu	Val	Val	Leu	Leu	Cys	Trp	Val	Lys	
				125					130					135	
Phe	Leu	Pro	Leu	Lys	Lys	Gln	Pro	Gly	Gln	Pro	Arg	Pro	Thr	Ser	
				140					145					150	
Lys	Pro	Pro	Ala	Ser	Gly	Ala	Ala	Ala	Asn	Val	Ser	Thr	Ser	Gly	
				155					160					165	
Ile	Thr	Pro	Gly	Gln	Ala	Ala	Ala	Ile	Ala	Ser	Thr	Thr	Ile	Met	
				170					175					180	
Val	Pro	Phe	Gly	Leu	Ile	Phe	Ile	Val	Phe	Ala	Val	His	Phe	Tyr	
				185					190					195	
Arg	Ser	Leu	Val	Ser	His	Lys	Thr	Asp	Arg	Gln	Phe	Gln	Glu	Leu	
				200					205					210	
Asn	Glu	Leu	Ala	Glu	Phe	Ala	Arg	Leu	Gln	Asp	Gln	Leu	Asp	His	
				215					220					225	
Arg	Gly	Asp	His	Pro	Leu	Thr	Pro	Gly	Ser	His	Tyr	Ala			
				230					235						

&lt;210&gt; 21

&lt;211&gt; 3298

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2847752CD1

**THIS PAGE BLANK (USPTO)**

```

<400> 21
Met Ala Arg Arg Pro Pro Trp Arg Gly Leu Gly Gly Arg Ser Thr
 1      5      10      15
Pro Ile Leu Leu Leu Leu Leu Leu Ser Leu Phe Pro Leu Ser Gln
 20      25      30
Glu Glu Leu Gly Gly Gly Gly His Gln Gly Trp Asp Pro Gly Leu
 35      40      45
Ala Ala Thr Thr Gly Pro Arg Ala His Ile Gly Gly Gly Ala Leu
 50      55      60
Ala Leu Cys Pro Glu Ser Ser Gly Val Arg Glu Asp Gly Gly Pro
 65      70      75
Gly Leu Gly Val Arg Glu Pro Ile Phe Val Gly Leu Arg Gly Arg
 80      85      90
Arg Gln Ser Ala Arg Asn Ser Arg Gly Pro Pro Glu Gln Pro Asn
 95      100     105
Glu Glu Leu Gly Ile Glu His Gly Val Gln Pro Leu Gly Ser Arg
 110     115     120
Glu Arg Glu Thr Gly Gln Gly Pro Gly Ser Val Leu Tyr Trp Arg
 125     130     135
Pro Glu Val Ser Ser Cys Gly Arg Thr Gly Pro Leu Gln Arg Gly
 140     145     150
Ser Leu Ser Pro Gly Ala Leu Ser Ser Gly Val Pro Gly Ser Gly
 155     160     165
Asn Ser Ser Pro Leu Pro Ser Asp Phe Leu Ile Arg His His Gly
 170     175     180
Pro Lys Pro Val Ser Ser Gln Arg Asn Ala Gly Thr Gly Ser Arg
 185     190     195
Lys Arg Val Gly Thr Ala Arg Cys Cys Gly Glu Leu Trp Ala Thr
 200     205     210
Gly Ser Lys Gly Gln Gly Glu Arg Ala Thr Thr Ser Gly Ala Glu
 215     220     225
Arg Thr Ala Pro Arg Arg Asn Cys Leu Pro Gly Ala Ser Gly Ser
 230     235     240
Gly Pro Glu Leu Asp Ser Ala Pro Arg Thr Ala Arg Thr Ala Pro
 245     250     255
Ala Ser Gly Ser Ala Pro Arg Glu Ser Arg Thr Ala Pro Glu Pro
 260     265     270
Ala Pro Lys Arg Met Arg Ser Arg Gly Leu Phe Arg Cys Arg Phe
 275     280     285
Leu Pro Gln Arg Pro Gly Pro Arg Pro Pro Gly Leu Pro Ala Arg
 290     295     300
Pro Glu Ala Arg Lys Val Thr Ser Ala Asn Arg Ala Arg Phe Arg
 305     310     315
Arg Ala Ala Asn Arg His Pro Gln Phe Pro Gln Tyr Asn Tyr Gln
 320     325     330
Thr Leu Val Pro Glu Asn Glu Ala Ala Gly Thr Ala Val Leu Arg
 335     340     345
Val Val Ala Gln Asp Pro Asp Ala Gly Glu Ala Gly Arg Leu Val
 350     355     360
Tyr Ser Leu Ala Ala Leu Met Asn Ser Arg Ser Leu Glu Leu Phe
 365     370     375
Ser Ile Asp Pro Gln Ser Gly Leu Ile Arg Thr Ala Ala Ala Leu
 380     385     390
Asp Arg Glu Ser Met Glu Arg His Tyr Leu Arg Val Thr Ala Gln
 395     400     405
Asp His Gly Ser Pro Arg Leu Ser Ala Thr Thr Met Val Ala Val
 410     415     420
Thr Val Ala Asp Arg Asn Asp His Ser Pro Val Phe Glu Gln Ala
 425     430     435
Gln Tyr Arg Glu Thr Leu Arg Glu Asn Val Glu Glu Gly Tyr Pro
 440     445     450
Ile Leu Gln Leu Arg Ala Thr Asp Gly Asp Ala Pro Pro Asn Ala
 455     460     465
Asn Leu Arg Tyr Arg Phe Val Gly Pro Pro Ala Ala Arg Ala Ala
 470     475     480
Ala Ala Ala Ala Phe Glu Ile Asp Pro Arg Ser Gly Leu Ile Ser
 485     490     495

```

**THIS PAGE BLANK (USPTO)**

Thr	Ser	Gly	Arg	Val	Asp	Arg	Glu	His	Met	Glu	Ser	Tyr	Glu	Leu
				500					505					510
Val	Val	Glu	Ala	Ser	Asp	Gln	Gly	Gln	Glu	Pro	Gly	Pro	Arg	Ser
				515					520					525
Ala	Thr	Val	Arg	Val	His	Ile	Thr	Val	Leu	Asp	Glu	Asn	Asp	Asn
				530					535					540
Ala	Pro	Gln	Phe	Ser	Glu	Lys	Arg	Tyr	Val	Ala	Gln	Val	Arg	Glu
				545					550					555
Asp	Val	Arg	Pro	His	Thr	Val	Val	Leu	Arg	Val	Thr	Ala	Thr	Asp
				560					565					570
Arg	Asp	Lys	Asp	Ala	Asn	Gly	Leu	Val	His	Tyr	Asn	Ile	Ile	Ser
				575					580					585
Gly	Asn	Ser	Arg	Gly	His	Phe	Ala	Ile	Asp	Ser	Leu	Thr	Gly	Glu
				590					595					600
Ile	Gln	Val	Val	Ala	Pro	Leu	Asp	Phe	Glu	Ala	Glu	Arg	Glu	Tyr
				605					610					615
Ala	Leu	Arg	Ile	Arg	Ala	Gln	Asp	Ala	Gly	Arg	Pro	Pro	Leu	Ser
				620					625					630
Asn	Asn	Thr	Gly	Leu	Ala	Ser	Ile	Gln	Val	Val	Asp	Ile	Asn	Asp
				635					640					645
His	Ile	Pro	Ile	Phe	Val	Ser	Thr	Pro	Phe	Gln	Val	Ser	Val	Leu
				650					655					660
Glu	Asn	Ala	Pro	Leu	Gly	His	Ser	Val	Ile	His	Ile	Gln	Ala	Val
				665					670					675
Asp	Ala	Asp	His	Gly	Glu	Asn	Ala	Arg	Leu	Glu	Tyr	Ser	Leu	Thr
				680					685					690
Gly	Val	Ala	Pro	Asp	Thr	Pro	Phe	Val	Ile	Asn	Ser	Ala	Thr	Gly
				695					700					705
Trp	Val	Ser	Val	Ser	Gly	Pro	Leu	Asp	Arg	Glu	Ser	Val	Glu	His
				710					715					720
Tyr	Phe	Phe	Gly	Val	Glu	Ala	Arg	Asp	His	Gly	Ser	Pro	Pro	Leu
				725					730					735
Ser	Ala	Ser	Ala	Ser	Val	Thr	Val	Thr	Val	Leu	Asp	Val	Asn	Asp
				740					745					750
Asn	Arg	Pro	Glu	Phe	Thr	Met	Lys	Glu	Tyr	His	Leu	Arg	Leu	Asn
				755					760					765
Glu	Asp	Ala	Ala	Val	Gly	Thr	Ser	Val	Val	Ser	Val	Thr	Ala	Val
				770					775					780
Asp	Arg	Asp	Ala	Asn	Ser	Ala	Ile	Ser	Tyr	Gln	Ile	Thr	Gly	Gly
				785					790					795
Asn	Thr	Arg	Asn	Arg	Phe	Ala	Ile	Ser	Thr	Gln	Gly	Gly	Val	Gly
				800					805					810
Leu	Val	Thr	Leu	Ala	Leu	Pro	Leu	Asp	Tyr	Lys	Gln	Glu	Arg	Tyr
				815					820					825
Phe	Lys	Leu	Val	Leu	Thr	Ala	Ser	Asp	Arg	Ala	Leu	His	Asp	His
				830					835					840
Cys	Tyr	Val	His	Ile	Asn	Ile	Thr	Asp	Ala	Asn	Thr	His	Arg	Pro
				845					850					855
Val	Phe	Gln	Ser	Ala	His	Tyr	Ser	Val	Ser	Val	Asn	Glu	Asp	Arg
				860					865					870
Pro	Met	Gly	Ser	Thr	Ile	Val	Val	Ile	Ser	Ala	Ser	Asp	Asp	Asp
				875					880					885
Val	Gly	Glu	Asn	Ala	Arg	Ile	Thr	Tyr	Leu	Leu	Glu	Asp	Asn	Leu
				890					895					900
Pro	Gln	Phe	Arg	Ile	Asp	Ala	Asp	Ser	Gly	Ala	Ile	Thr	Leu	Gln
				905					910					915
Ala	Pro	Leu	Asp	Tyr	Glu	Asp	Gln	Val	Thr	Tyr	Thr	Leu	Ala	Ile
				920					925					930
Thr	Ala	Arg	Asp	Asn	Gly	Ile	Pro	Gln	Lys	Ala	Asp	Thr	Thr	Tyr
				935					940					945
Val	Glu	Val	Met	Val	Asn	Asp	Val	Asn	Asp	Asn	Ala	Pro	Gln	Phe
				950					955					960
Val	Ala	Ser	His	Tyr	Thr	Gly	Leu	Val	Ser	Glu	Asp	Ala	Pro	Pro
				965					970					975
Phe	Thr	Ser	Val	Leu	Gln	Ile	Ser	Ala	Thr	Asp	Arg	Asp	Ala	His
				980					985					990
Ala	Asn	Gly	Arg	Val	Gln	Tyr	Thr	Phe	Gln	Asn	Gly	Glu	Asp	Gly

**THIS PAGE BLANK (USPTO)**



Asp Gly Asp Phe Thr	Ile Glu Pro Thr Ser	Gly Ile Val Arg Thr	995	1000	1005
Val Arg Arg Leu Asp	Arg Glu Ala Val Ser	Val Tyr Glu Leu Thr	1010	1015	1020
Ala Tyr Ala Val Asp	Arg Gly Val Pro Pro	Leu Arg Thr Pro Val	1025	1030	1035
Ser Ile Gln Val Met	Val Gln Asp Val Asn	Asp Asn Ala Pro Val	1040	1045	1050
Phe Pro Ala Glu Glu	Phe Glu Val Arg Val	Lys Glu Asn Ser Ile	1055	1060	1065
Val Gly Ser Val Val	Ala Gln Ile Thr Ala	Val Asp Pro Asp Glu	1070	1075	1080
Gly Pro Asn Ala His	Ile Met Tyr Gln Ile	Val Glu Gly Asn Ile	1085	1090	1095
Pro Glu Leu Phe Gln	Met Asp Ile Phe Ser	Gly Glu Leu Thr Ala	1100	1105	1110
Leu Ile Asp Leu Asp	Tyr Glu Ala Arg Gln	Glu Tyr Val Ile Val	1115	1120	1125
Val Gln Ala Thr Ser	Ala Pro Leu Val Ser	Arg Ala Thr Val His	1130	1135	1140
Val Arg Leu Val Asp	Gln Asn Asp Asn Ser	Pro Val Leu Asn Asn	1145	1150	1155
Phe Gln Ile Leu Phe	Asn Asn Tyr Val Ser	Asn Arg Ser Asp Thr	1160	1165	1170
Phe Pro Ser Gly Ile	Ile Gly Arg Ile Pro	Ala Tyr Asp Pro Asp	1175	1180	1185
Val Ser Asp His Leu	Phe Tyr Ser Phe Glu	Arg Gly Asn Glu Leu	1190	1195	1200
Gln Leu Leu Val Val	Asn Gln Thr Ser Gly	Glu Leu Arg Leu Ser	1205	1210	1215
Arg Lys Leu Asp Asn	Asn Arg Pro Leu Val	Ala Ser Met Leu Val	1220	1225	1230
Thr Val Thr Asp Gly	Leu His Ser Val Thr	Ala Gln Cys Val Leu	1235	1240	1245
Arg Val Val Ile Ile	Thr Glu Glu Leu Leu	Ala Asn Ser Leu Thr	1250	1255	1260
Val Arg Leu Glu Asn	Met Trp Gln Glu Arg	Phe Leu Ser Pro Leu	1265	1270	1275
Leu Gly Arg Phe Leu	Glu Gly Val Ala Ala	Val Leu Ala Thr Pro	1280	1285	1290
Ala Glu Asp Val Phe	Ile Phe Asn Ile Gln	Asn Asp Thr Asp Val	1295	1300	1305
Gly Gly Thr Val Leu	Asn Val Ser Phe Ser	Ala Leu Ala Pro Arg	1310	1315	1320
Gly Ala Gly Ala Gly	Ala Ala Gly Pro Trp	Phe Ser Ser Glu Glu	1325	1330	1335
Leu Gln Glu Gln Leu	Tyr Val Arg Arg Ala	Ala Leu Ala Ala Arg	1340	1345	1350
Ser Leu Leu Asp Val	Leu Pro Phe Asp Asp	Asn Val Cys Leu Arg	1355	1360	1365
Glu Pro Cys Glu Asn	Tyr Met Lys Cys Val	Ser Val Leu Arg Phe	1370	1375	1380
Asp Ser Ser Ala Pro	Phe Leu Ala Ser Ala	Ser Thr Leu Phe Arg	1385	1390	1395
Pro Ile Gln Pro Ile	Ala Gly Leu Arg Cys	Arg Cys Pro Pro Gly	1400	1405	1410
Phe Thr Gly Asp Phe	Cys Glu Thr Glu Leu	Asp Leu Cys Tyr Ser	1415	1420	1425
Asn Pro Cys Arg Asn	Gly Gly Ala Cys Ala	Arg Arg Glu Gly Gly	1430	1435	1440
Tyr Thr Cys Val Cys	Arg Pro Arg Phe Thr	Gly Glu Asp Cys Glu	1445	1450	1455
Leu Asp Thr Glu Ala	Gly Arg Cys Val Pro	Gly Val Cys Arg Asn	1460	1465	1470
Gly Gly Thr Cys Thr	Asp Ala Pro Asn Gly	Gly Phe Arg Cys Gln	1475	1480	1485
			1490	1495	1500

**THIS PAGE BLANK (USPTO)**

Cys	Pro	Ala	Gly	Gly	Ala	Phe	Glu	Gly	Pro	Arg	Cys	Glu	Val	Ala
			1505						1510					1515
Ala	Arg	Ser	Phe	Pro	Pro	Ser	Ser	Phe	Val	Met	Phe	Arg	Gly	Leu
			1520						1525					1530
Arg	Gln	Arg	Phe	His	Leu	Thr	Leu	Ser	Leu	Ser	Phe	Ala	Thr	Val
			1535						1540					1545
Gln	Gln	Ser	Gly	Leu	Leu	Phe	Tyr	Asn	Gly	Arg	Leu	Asn	Glu	Lys
			1550						1555					1560
His	Asp	Phe	Leu	Ala	Leu	Glu	Leu	Val	Ala	Gly	Gln	Val	Arg	Leu
			1565						1570					1575
Thr	Tyr	Ser	Thr	Gly	Glu	Ser	Asn	Thr	Val	Val	Ser	Pro	Thr	Val
			1580						1585					1590
Pro	Gly	Gly	Leu	Ser	Asp	Gly	Gln	Trp	His	Thr	Val	His	Leu	Arg
			1595						1600					1605
Tyr	Tyr	Asn	Lys	Pro	Arg	Thr	Asp	Ala	Leu	Gly	Gly	Ala	Gln	Gly
			1610						1615					1620
Pro	Ser	Lys	Asp	Lys	Val	Ala	Val	Leu	Ser	Val	Asp	Asp	Cys	Asp
			1625						1630					1635
Val	Ala	Val	Ala	Leu	Gln	Phe	Gly	Ala	Glu	Ile	Gly	Asn	Tyr	Ser
			1640						1645					1650
Cys	Ala	Ala	Ala	Gly	Val	Gln	Thr	Ser	Ser	Lys	Lys	Ser	Leu	Asp
			1655						1660					1665
Leu	Thr	Gly	Pro	Leu	Leu	Leu	Gly	Gly	Val	Pro	Asn	Leu	Pro	Glu
			1670						1675					1680
Asn	Phe	Pro	Val	Ser	His	Lys	Asp	Phe	Ile	Gly	Cys	Met	Arg	Asp
			1685						1690					1695
Leu	His	Ile	Asp	Gly	Arg	Arg	Val	Asp	Met	Ala	Ala	Phe	Val	Ala
			1700						1705					1710
Asn	Asn	Gly	Thr	Met	Ala	Gly	Cys	Gln	Ala	Lys	Leu	His	Phe	Cys
			1715						1720					1725
Asp	Ser	Gly	Pro	Cys	Lys	Asn	Ser	Gly	Phe	Cys	Ser	Glu	Arg	Trp
			1730						1735					1740
Gly	Ser	Phe	Ser	Cys	Asp	Cys	Pro	Val	Gly	Phe	Gly	Gly	Lys	Asp
			1745						1750					1755
Cys	Gln	Leu	Thr	Met	Ala	His	Pro	His	His	Phe	Arg	Gly	Asn	Gly
			1760						1765					1770
Thr	Leu	Ser	Trp	Asn	Phe	Gly	Ser	Asp	Met	Ala	Val	Ser	Val	Pro
			1775						1780					1785
Trp	Tyr	Leu	Gly	Leu	Ala	Phe	Arg	Thr	Arg	Ala	Thr	Gln	Gly	Val
			1790						1795					1800
Leu	Met	Gln	Val	Gln	Ala	Gly	Pro	His	Ser	Thr	Leu	Leu	Cys	Gln
			1805						1810					1815
Leu	Asp	Arg	Gly	Leu	Leu	Ser	Val	Thr	Val	Thr	Arg	Gly	Ser	Gly
			1820						1825					1830
Arg	Ala	Ser	His	Leu	Leu	Leu	Asp	Gln	Val	Thr	Val	Ser	Asp	Gly
			1835						1840					1845
Arg	Trp	His	Asp	Leu	Arg	Leu	Glu	Leu	Gln	Glu	Glu	Pro	Gly	Gly
			1850						1855					1860
Arg	Arg	Gly	His	His	Val	Leu	Met	Val	Ser	Leu	Asp	Phe	Ser	Leu
			1865						1870					1875
Phe	Gln	Asp	Thr	Met	Ala	Val	Gly	Ser	Glu	Leu	Gln	Gly	Leu	Lys
			1880						1885					1890
Val	Lys	Gln	Leu	His	Val	Gly	Gly	Leu	Pro	Pro	Gly	Ser	Ala	Glu
			1895						1900					1905
Glu	Ala	Pro	Gln	Gly	Leu	Val	Gly	Cys	Ile	Gln	Pro	Pro	Ser	Glu
			1910						1915					1920
Cys	Gly	Pro	Gly	Cys	Val	Val	Thr	Asn	Ala	Cys	Ala	Ser	Gly	Pro
			1925						1930					1935
Cys	Pro	Pro	His	Ala	Asp	Cys	Arg	Asp	Leu	Trp	Gln	Thr	Phe	Ser
			1940						1945					1950
Cys	Thr	Cys	Gln	Pro	Gly	Tyr	Tyr	Gly	Pro	Gly	Cys	Val	Asp	Ala
			1955						1960					1965
Cys	Leu	Leu	Asn	Pro	Cys	Gln	Asn	Gln	Gly	Ser	Cys	Arg	His	Leu
			1970						1975					1980
Pro	Gly	Ala	Pro	His	Gly	Tyr	Thr	Cys	Asp	Cys	Val	Gly	Gly	Tyr
			1985						1990					1995
Phe	Gly	His	His	Cys	Glu	His	Arg	Met	Asp	Gln	Gln	Cys	Pro	Arg

**THIS PAGE BLANK (USPTO)**

	2000		2005		2010
Gly Trp Trp Gly Ser Pro Thr Cys Gly Pro Cys Asn Cys Asp Val					
	2015		2020		2025
His Lys Gly Phe Asp Pro Asn Cys Asn Lys Thr Asn Gly Gln Cys					
	2030		2035		2040
His Cys Lys Glu Phe His Tyr Arg Pro Arg Gly Ser Asp Ser Cys					
	2045		2050		2055
Leu Pro Cys Asp Cys Tyr Pro Val Gly Ser Thr Ser Arg Ser Cys					
	2060		2065		2070
Ala Pro His Ser Gly Gln Cys Pro Cys Arg Pro Gly Ala Leu Gly					
	2075		2080		2085
Arg Gln Cys Asn Ser Cys Asp Ser Pro Phe Ala Glu Val Thr Ala					
	2090		2095		2100
Ser Gly Cys Arg Val Leu Tyr Asp Ala Cys Pro Lys Ser Leu Arg					
	2105		2110		2115
Ser Gly Val Trp Trp Pro Gln Thr Lys Phe Gly Val Leu Ala Thr					
	2120		2125		2130
Val Pro Cys Pro Arg Gly Ala Leu Gly Leu Arg Gly Ala Gly Ala					
	2135		2140		2145
Ala Val Arg Leu Cys Asp Glu Ala Gln Gly Trp Leu Glu Pro Asp					
	2150		2155		2160
Leu Phe Asn Cys Thr Ser Pro Ala Phe Arg Glu Leu Ser Leu Leu					
	2165		2170		2175
Leu Asp Gly Leu Glu Leu Asn Lys Thr Ala Leu Asp Thr Met Glu					
	2180		2185		2190
Ala Lys Lys Leu Ala Gln Arg Leu Arg Glu Val Thr Gly His Thr					
	2195		2200		2205
Asp His Tyr Phe Ser Gln Asp Val Arg Val Thr Ala Arg Leu Leu					
	2210		2215		2220
Ala His Leu Leu Ala Phe Glu Ser His Gln Gln Gly Phe Gly Leu					
	2225		2230		2235
Thr Ala Thr Gln Asp Ala His Phe Asn Glu Asn Leu Leu Trp Ala					
	2240		2245		2250
Gly Ser Ala Leu Leu Ala Pro Glu Thr Gly Asp Leu Trp Ala Ala					
	2255		2260		2265
Leu Gly Gln Arg Ala Pro Gly Gly Ser Pro Gly Ser Ala Gly Leu					
	2270		2275		2280
Val Arg His Leu Glu Glu Tyr Ala Ala Thr Leu Ala Arg Asn Met					
	2285		2290		2295
Glu Leu Thr Tyr Leu Asn Pro Met Gly Leu Val Thr Pro Asn Ile					
	2300		2305		2310
Met Leu Ser Ile Asp Arg Met Glu His Pro Ser Ser Pro Arg Gly					
	2315		2320		2325
Ala Arg Arg Tyr Pro Arg Tyr His Ser Asn Leu Phe Arg Gly Gln					
	2330		2335		2340
Asp Ala Trp Asp Pro His Thr His Val Leu Leu Pro Ser Gln Ser					
	2345		2350		2355
Pro Arg Pro Ser Pro Ser Glu Val Leu Pro Thr Ser Ser Ser Ile					
	2360		2365		2370
Glu Asn Ser Thr Thr Ser Ser Val Val Pro Pro Pro Ala Pro Pro					
	2375		2380		2385
Glu Pro Glu Pro Gly Ile Ser Ile Ile Ile Leu Leu Val Tyr Arg					
	2390		2395		2400
Thr Leu Gly Gly Leu Leu Pro Ala Gln Phe Gln Ala Glu Arg Arg					
	2405		2410		2415
Gly Ala Arg Leu Pro Gln Asn Pro Val Met Asn Ser Pro Val Val					
	2420		2425		2430
Ser Val Ala Val Phe His Gly Arg Asn Phe Leu Arg Gly Ile Leu					
	2435		2440		2445
Glu Ser Pro Ile Ser Leu Glu Phe Arg Leu Leu Gln Thr Ala Asn					
	2450		2455		2460
Arg Ser Lys Ala Ile Cys Val Gln Trp Asp Pro Pro Gly Leu Ala					
	2465		2470		2475
Glu Gln His Gly Val Trp Thr Ala Arg Asp Cys Glu Leu Val His					
	2480		2485		2490
Arg Asn Gly Ser His Ala Arg Cys Arg Cys Ser Arg Thr Gly Thr					
	2495		2500		2505

**THIS PAGE BLANK (USPTO)**

Phe	Gly	Val	Leu	Met	Asp	Ala	Ser	Pro	Arg	Glu	Arg	Leu	Glu	Gly
				2510					2515					2520
Asp	Leu	Glu	Leu	Leu	Ala	Val	Phe	Thr	His	Val	Val	Val	Ala	Val
				2525					2530					2535
Ser	Val	Ala	Ala	Leu	Val	Leu	Thr	Ala	Ala	Ile	Leu	Leu	Ser	Leu
				2540					2545					2550
Arg	Ser	Leu	Lys	Ser	Asn	Val	Arg	Gly	Ile	His	Ala	Asn	Val	Ala
				2555					2560					2565
Ala	Ala	Leu	Gly	Val	Ala	Glu	Leu	Leu	Phe	Leu	Leu	Gly	Ile	His
				2570					2575					2580
Arg	Thr	His	Asn	Gln	Leu	Val	Cys	Thr	Ala	Val	Ala	Ile	Leu	Leu
				2585					2590					2595
His	Tyr	Phe	Phe	Leu	Ser	Thr	Phe	Ala	Trp	Leu	Phe	Val	Gln	Gly
				2600					2605					2610
Leu	His	Leu	Tyr	Arg	Met	Gln	Val	Glu	Pro	Arg	Asn	Val	Asp	Arg
				2615					2620					2625
Gly	Ala	Met	Arg	Phe	Tyr	His	Ala	Leu	Gly	Trp	Gly	Val	Pro	Ala
				2630					2635					2640
Val	Leu	Leu	Gly	Leu	Ala	Val	Gly	Leu	Asp	Pro	Glu	Gly	Tyr	Gly
				2645					2650					2655
Asn	Pro	Asp	Phe	Cys	Trp	Ile	Ser	Val	His	Glu	Pro	Leu	Ile	Trp
				2660					2665					2670
Ser	Phe	Ala	Gly	Pro	Val	Val	Leu	Val	Ile	Val	Met	Asn	Gly	Thr
				2675					2680					2685
Met	Phe	Leu	Leu	Ala	Ala	Arg	Thr	Ser	Cys	Ser	Thr	Gly	Gln	Arg
				2690					2695					2700
Glu	Ala	Lys	Lys	Thr	Ser	Ala	Leu	Thr	Leu	Arg	Ser	Ser	Phe	Leu
				2705					2710					2715
Leu	Leu	Leu	Leu	Val	Ser	Ala	Ser	Trp	Leu	Phe	Gly	Leu	Leu	Ala
				2720					2725					2730
Val	Asn	His	Ser	Ile	Leu	Ala	Phe	His	Tyr	Leu	His	Ala	Gly	Leu
				2735					2740					2745
Cys	Gly	Leu	Gln	Gly	Leu	Ala	Val	Leu	Leu	Leu	Phe	Cys	Val	Leu
				2750					2755					2760
Asn	Ala	Asp	Ala	Arg	Ala	Ala	Trp	Met	Pro	Ala	Cys	Leu	Gly	Arg
				2765					2770					2775
Lys	Ala	Ala	Pro	Glu	Glu	Ala	Arg	Pro	Ala	Pro	Gly	Leu	Gly	Pro
				2780					2785					2790
Gly	Ala	Tyr	Asn	Asn	Thr	Ala	Leu	Phe	Glu	Glu	Ser	Gly	Leu	Ile
				2795					2800					2805
Arg	Ile	Thr	Leu	Gly	Ala	Ser	Thr	Val	Ser	Ser	Val	Ser	Ser	Ala
				2810					2815					2820
Arg	Ser	Gly	Arg	Thr	Gln	Asp	Gln	Asp	Ser	Gln	Arg	Gly	Arg	Ser
				2825					2830					2835
Tyr	Leu	Arg	Asp	Asn	Val	Leu	Val	Arg	His	Gly	Ser	Ala	Ala	Asp
				2840					2845					2850
His	Thr	Asp	His	Ser	Leu	Gln	Ala	His	Ala	Gly	Pro	Thr	Asp	Leu
				2855					2860					2865
Asp	Val	Ala	Met	Phe	His	Arg	Asp	Ala	Gly	Ala	Asp	Ser	Asp	Ser
				2870					2875					2880
Asp	Ser	Asp	Leu	Ser	Leu	Glu	Glu	Glu	Arg	Ser	Leu	Ser	Ile	Pro
				2885					2890					2895
Ser	Ser	Glu	Ser	Glu	Asp	Asn	Gly	Arg	Thr	Arg	Gly	Arg	Val	Gln
				2900					2905					2910
Arg	Pro	Leu	Cys	Arg	Ala	Ala	Gln	Ser	Glu	Arg	Leu	Leu	Thr	His
				2915					2920					2925
Pro	Lys	Asp	Val	Asp	Gly	Asn	Asp	Leu	Leu	Ser	Tyr	Trp	Pro	Ala
				2930					2935					2940
Leu	Gly	Glu	Cys	Glu	Ala	Ala	Pro	Cys	Ala	Leu	Gln	Thr	Trp	Gly
				2945					2950					2955
Ser	Glu	Arg	Arg	Leu	Gly	Leu	Asp	Thr	Ser	Lys	Asp	Ala	Ala	Asn
				2960					2965					2970
Asn	Asn	Gln	Pro	Asp	Pro	Ala	Leu	Thr	Ser	Gly	Asp	Glu	Thr	Ser
				2975					2980					2985
Leu	Gly	Arg	Ala	Gln	Arg	Gln	Arg	Lys	Gly	Ile	Leu	Lys	Asn	Arg
				2990					2995					3000
Leu	Gln	Tyr	Pro	Leu	Val	Pro	Gln	Thr	Arg	Gly	Ala	Pro	Glu	Leu

**THIS PAGE BLANK (USPTO)**



3005	3010	3015
Ser Trp Cys Arg Ala	Ala Thr Leu Gly His	Arg Ala Val Pro Ala
3020	3025	3030
Ala Ser Tyr Gly Arg	Ile Tyr Ala Gly Gly	Gly Thr Gly Ser Leu
3035	3040	3045
Ser Gln Pro Ala Ser	Arg Tyr Ser Ser Arg	Glu Gln Leu Asp Leu
3050	3055	3060
Leu Leu Arg Arg Gln	Leu Ser Arg Glu Arg	Leu Glu Glu Ala Pro
3065	3070	3075
Ala Pro Val Leu Arg	Pro Leu Ser Arg Pro	Gly Ser Gln Glu Cys
3080	3085	3090
Met Asp Ala Ala Pro	Gly Arg Leu Glu Pro	Lys Asp Arg Gly Ser
3095	3100	3105
Thr Leu Pro Arg Arg	Gln Pro Pro Arg Asp	Tyr Pro Gly Ala Met
3110	3115	3120
Ala Gly Arg Phe Gly	Ser Arg Asp Ala Leu	Asp Leu Gly Ala Pro
3125	3130	3135
Arg Glu Trp Leu Ser	Thr Leu Pro Pro Pro	Arg Arg Thr Arg Asp
3140	3145	3150
Leu Asp Pro Gln Pro	Pro Pro Leu Pro Leu	Ser Pro Gln Arg Gln
3155	3160	3165
Leu Ser Arg Asp Pro	Leu Leu Pro Ser Arg	Pro Leu Asp Ser Leu
3170	3175	3180
Ser Arg Ser Ser Asn	Ser Arg Glu Gln Leu	Asp Gln Val Pro Ser
3185	3190	3195
Arg His Pro Ser Arg	Glu Ala Leu Gly Pro	Leu Pro Gln Leu Leu
3200	3205	3210
Arg Ala Arg Glu Asp	Ser Val Ser Gly Pro	Ser His Gly Pro Ser
3215	3220	3225
Thr Glu Gln Leu Asp	Ile Leu Ser Ser Ile	Leu Ala Ser Phe Asn
3230	3235	3240
Ser Ser Ala Leu Ser	Ser Val Gln Ser Ser	Ser Thr Pro Leu Gly
3245	3250	3255
Pro His Thr Thr Ala	Thr Pro Ser Ala Thr	Ala Ser Val Leu Gly
3260	3265	3270
Pro Ser Thr Pro Arg	Ser Ala Thr Ser His	Ser Ile Ser Glu Leu
3275	3280	3285
Ser Pro Asp Ser Glu	Val Pro Arg Ser Glu	Gly His Ser
3290	3295	

&lt;210&gt; 22

&lt;211&gt; 905

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1424691CB1

&lt;400&gt; 22

```

cccacgcgtc cgtgttagaa ctccaaaaag cactaaacat gtccattcga ttcagtggag 60
tagaacaataa cctcaggatg aagtgaagc agtccaaact gccattcaga cattattcac 120
caattcagat ggcaaccctg gaagcaggtc cgactcaagt gctgattgcc agtgggttaga 180
tactctgagg atgcggcaga ttgcatccaa cacttcttta cagcgttccc agagcaatcc 240
tattctgggg tcaccgttct tctcacactt tgatggccag gattcctacg ctgctgctgt 300
gagacggccc caggtgcccc ttaagtatca acagattaca cctgtgaacc agtccagaag 360
ctcgtctcct actcagtatg gactgaccaa aaacttctct tccctacatc tcaactctag 420
ggacagtggc ttttccagtg gcaataactga cacctcttca gagagggggtc gatactcaga 480
cagaagcagg aacaaatagc gacgtggtag tatatcactc aattcttctc ctagaggaag 540
atacagtggg aagagtcagc attccactcc ttcaagagga agataccctg gaaagttcta 600
cagggtttct cagtcagcac tcaatcctca ccagtcgcct gacttcaaga gaagccccag 660
ggacctccac caaccaaca ccataccagg gatgcctttg caccctgaga ctgactcaag 720
agccagtga gaggacagca aagtcagcga agggggctgg acaaaagtgg aataccggaa 780
aaagccccc aggcctctc cgcgcaaac caataaagag agagccagag gggaccaccg 840
tggatggaga aacttttgat gaattgaact acatagcttt tctaagcagg ttaaaaaaaa 900
aaaaa                                           905

```

**THIS PAGE BLANK (USPTO)**

<210> 23  
<211> 2390  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 1450801CB1

<400> 23  
tgaccttggc accccgcttc ctgccaaacg ccctgatcag tgtggacttt gctgccaaact 60  
atctcaccaa gatctatggg ctcacctttg gccagaagcc aaacttgagg tctgtgtacc 120  
tgcacaacaa caagctggca gacgcccggc tgccggacaa catgttcaac ggctccagca 180  
acgtcgagggt cctcatcctg tccagcaact tcttgcgcca cgtgcccagc cacctgccgc 240  
ctgccctgta caagctgcac ctcaagaaca acaagctgga gaagatcccc ccggggggcct 300  
tgcagcgagc tggagcagcc tgcgcgagct atacctgcag aacaactacc tgactgacga 360  
gggcctggac aacgagacct tctggaagct ctccagcctg gactacctgg atctgtccag 420  
caacaacctg tctcgggtcc cagctgggct gccgcgcagc ctggtgctgc tgcacttggg 480  
gaagaacgcc atccggagcg tggagcgcaa tgtgtgacc cccatccgca gcctggagta 540  
cctgctcagc cacagcaacc agctgcggga gcagctgcac caccactgg cctccagggg 600  
cctcaagcgg ttgcacacgg tgcacctgta caacaacgg ctggagcgcg tgcccagtg 660  
cctgcctcgc cgcgtgcgca cctcatgat cctgcacaa cagatcacag gcattggccg 720  
cgaagacttt gccaccacct acttctctga ggagctcaac ctgagctaca accgcatcac 780  
cagggacatg gtgcaccgca acgcttccg catcctgcgc ctgctgcgt cgtgggact 840  
gtcgggcaac cggctgcaca cgtctgccac tgggctgcct cgaaatgtcc atgtgtgaa 900  
gggtcaagcgc aatgagctgg ctgccttggc acgagggggc ctggcgggca tggctcagct 960  
gcgtgagctg tacctcacca gcaaccgact gcgcagccga gccctggggc cccgtgcctg 1020  
gggtggacat gcccatctgc agctgctgga catcgcggg aatcagctca cagagatccc 1080  
cgaggggctc cccgagtcac ttgagtacct gtacctgcag aacaacaaga ttagtgcggt 1140  
gcccgcgaat gccttcgact ccacgcccac cctcaagggg atctttctca ggtttaacaa 1200  
gctggctgtg ggctccgtgg tggacagtgc cttccggagg ctgaagcacc tgcaggctct 1260  
ggacattgaa ggcaacttag agtttgggta catttccaag gaccgtggcc gcttggggaa 1320  
ggaaaaggag gaggaggaag agggaggagg ggaggaagag gaaacaagat agtgacaagg 1380  
tgatgcagat gtgacctagg atgatggacc gccggactct tttctgcagc acacgcctgt 1440  
gtgctgtgag ccccccactc tgccgtgtct acacagacac acccagctgc acacatgagg 1500  
catcccacat gacacgggct gacacagtct catatcccca ccccttcca cggcgtgtcc 1560  
cacggccaga cacatgcaca cacatcacac cctcaaacac ccagctcagc cacacacaac 1620  
taccctccaa accaccacag tctctgtcac acccccacta ccgctgccac gccctctgaa 1680  
tcatgcaggg aagggtctgc ccttgccctg gcacacacag gcacccatc cctccccctg 1740  
ctgacatgtg tatgcgtatg catacacacc acacacacac acatgcacaa gtcattgtcg 1800  
aacagccctc caaagcctat gccacagaca gctcttgccc cagccagaat cagccatagc 1860  
agctcgccgt ctgccctgtc catctgtccg tccgttccct ggagaagaca caagggtatc 1920  
catgctctgt ggccagggtgc ctgccaccct ctggaactca caaaagctgg cttttattcc 1980  
tttcccaccc tatggggaca ggagccttca ggactgctgg cctggcctgg ccacccctgc 2040  
tctccagggt gctgggcagt cactctgtca agagtccctc cctggcacgc cctggcagga 2100  
cacaggcact ttccaatgg gcaagcccag tggaggcagg atgggagagc cccctgggtg 2160  
ctgctggggc cttggggcag gagtgaagca gaggtgatgg ggcctgggctg agccagggag 2220  
gaaggaccca gctgcaccta ggagacacct ttgttcttca ggcctgtggg ggaagttccg 2280  
gggtgccttta ttttttattc ttttctaagg aaaaaaatga taaaaatctc aaagctgatt 2340  
tttcttggtta tagaaaaact aatataaaag cattatccct aaaaaaaaaa 2390

<210> 24  
<211> 1471  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 1597872CB1

<400> 24  
cttggctagg agcccgtgc gggccaggcg cgcggtgcgc aggcataccc gggctcgggg 60  
tgaacaccgc ggagccccgg gacttgccct tgcgttgcctg tggcggcatg gtcttctgtc 120  
tgtcgagcga ggagccgcgc cgcgcgctgc gaacgacatg gtccacttcc aggcctcgga 180  
agtccagcag ctgctacaca acaagttcgt ggtcatcttg ggggactcca ttcagagggc 240  
tgtgtacaag gacctggtgc tcttgctcca gaaagactca ctgctcacag ctgcccagct 300  
gaaagccaag taccttgagg atgttctgga agagctgaca tatggacctg ccccggaact 360

**THIS PAGE BLANK (USPTO)**

ggtgatcatc	aactcctgcc	tctgggatct	ctccagatat	ggtcgctgct	caatggagag	420
ctaccgggag	aacctggagc	gggtgtttgt	gcgcattggac	caagtattgc	cagactcctg	480
cctgctggtg	tggaacatgg	cgatgcccc	cggggaacgt	atcactgggg	gtttcctcct	540
gccagagctc	cagccccctg	caggctccct	gcggcgggat	gtggttgaag	ggaacttcta	600
cagtgtctac	ctggccgggg	accactgctt	tgatgtccta	gacctccact	ttcacttccg	660
gcatgcagta	cagcaccgtc	atcgggatgg	tgctccactgg	gaccagcatg	cacaccgcca	720
cctctcacac	ctgcttctga	cccatgtggc	tgacgcctgg	ggcgtggagc	tgcccaagcg	780
tggctatccc	cctgaaccgt	ggattgagga	ctgggcagag	atgaatcate	cattccaggg	840
aagccatagg	cagaccccag	acttcgggga	gcacctggcc	ttgctccac	ccccaccttc	900
ttctttgcct	cctcccatgc	cttttcccta	cccgttctct	cagccccctg	cacctccctc	960
cttcccaccc	ctgccccagg	ataccccctt	tttcccaggc	cagccccctc	caccccatga	1020
attcttcaac	tataatccag	tggaggactt	ctgagtgcca	ccccacttag	gatgtggccc	1080
tggagtgaac	tttgtgcctg	gcccctctgc	acctccaatc	cctggcccta	atccccatgg	1140
tcagcactgg	ggcccagtg	tccaccgggg	gatgccacgc	tatgttcccta	acagccccta	1200
ccatgtgcgg	agaaatgggg	ggccctgcag	gcagcggctc	agacactcag	agagactgat	1260
ccacacatac	aaactgggaca	gacggcctcc	tgcccatctg	gggacatggc	ctgggtgac	1320
tggatcttgg	gctgggactg	gatgtgccaa	tgcccttca	gggcctgct	ggcacctcag	1380
gtactgggct	aggggtgtctg	ctatgcctgg	tattgttctt	gtccattgct	gtcaccaata	1440
aaggcatgga	agaacagagt	gaaaaaaaaa	a			1471

&lt;210&gt; 25

&lt;211&gt; 3293

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1674661CB1

&lt;400&gt; 25

gcgagcgcg	gcaggcggg	acgcggggg	aggggtggac	ggcggtcaga	gccgaacg	60
agggcgggc	ccggggactg	gagctgcgc	caataggaca	gctggcctga	agctcagag	120
cggggcggtg	gccatggccc	cacactgggc	tgtctggctg	ctggcagcaa	ggctgtggg	180
cctgggcat	ggggctgagg	tgtgggtgaa	ccttgtgccg	cgtaagacag	tgtcttctg	240
ggagctggcc	acggtagtac	ggcggttctc	ccagaccggc	atccaggact	tcttgacact	300
gacgtgacg	gagcccactg	ggcttctgta	cgtgggcgcc	cgagaggccc	tgtttggctt	360
cagcatggag	gccctggagc	tgcaaggagc	gatctcctgg	gaggcccccg	tggaagaaga	420
gactgagtgt	atccagaaag	ggaagaacaa	ccagaccgag	tgcttcaact	tcatccgctt	480
cctgcagccc	tacaatgcct	cccacctgta	cgtctgtggc	acctacgcct	tccagcccaa	540
gtgcacatc	tgcaacatgc	tcaccttcac	tttggagcat	ggagagtttg	aagatgggaa	600
gggcaagtgt	ccctatgacc	cagctaaggg	ccatgctggc	cttcttgtgg	atgggtgagct	660
gtactcgggc	acactcaaca	acttctctgg	cacggaaccc	attatcctgc	gtaacatggg	720
gccccaccac	tccatgaaga	cagagtacct	ggccttttgg	ctcaacgaac	ctcactttgt	780
aggetctg	tattgtacct	agactgtggg	cttcttcacg	ggggacgacg	acaaggctta	840
cttcttcttc	agggagcggg	cactggagtc	cgactgctat	gccgagcagg	tgggtggctg	900
tgtggcccg	gtctgcaagg	gcgatatggg	gggcgcacgg	accctgcaga	ggaagtggac	960
cagtttctg	aaggcgcggc	tggcatgctc	tgccccgaac	tggcagctct	acttcaacca	1020
gctgcaggcg	atgcacaccc	tcaggacac	ctcctggcac	aacaccacct	tctttgggg	1080
ttttcaagca	cagtgggggtg	acatgtacct	gtcggccatc	tgtgagtacc	agttggaaga	1140
gatccagcgg	gtgtttgagg	gccccataaa	ggagtaccat	gaggaagccc	agaagtggga	1200
ccgtacact	gacctgttac	ccagccctcg	gcttggctcg	tgcatataca	actggcatcg	1260
gcgccacggc	taccacagct	ccctggagct	accgacaac	atcctcaact	tcgtcaagaa	1320
gcacccgctg	atggaggagc	agggtggggc	tcgggtggagc	cgccccctgc	tcgtgaagaa	1380
gggcaccaac	ttcaccaccc	tgggtggccga	cgggtttaca	ggacttgatg	gagccacct	1440
tacagtgtg	ttcattggca	caggagacgg	ctggctgctc	aaggctgtga	gcctggggcc	1500
ctgggttcac	ctgattgagg	agctgcagct	gtttgaccag	gagcccatga	gaagcctgg	1560
gctatctcag	agcaagaagc	tgtcttttgc	cggctcccg	tctcagctgg	tcagctgct	1620
cgtggccgac	tgcatgaagt	atcgctcctg	tcagacatgt	gtcctcgccc	gggaccccta	1680
ttgcgcctgg	agcgtcaaca	ccagccgctg	tgtggccctg	ggtggccact	ctggatctct	1740
actgatccag	catgtgatga	cctcggaac	ttcaggcatc	tgcaacctcc	gtggcagtaa	1800
gaaagtccag	ccactccca	aaaacatcac	ggtggtggcg	ggcacagacc	tgggtgctgc	1860
ctgccacctc	tcttccaact	tgccccatgc	ccgctggacc	tttggggggc	gggacctg	1920
tgccgaacag	ccgggtctct	tctctacga	tgccccgctc	caggccctgg	ttgtgatggc	1980
tgccacgccc	cgccatgccg	ggccctacca	ctgcttttca	gaggagcagg	gggcgcggct	2040
ggctgctgaa	ggctaccttg	tggctgtcgt	ggcaggcccg	tcggtgacct	tggaggcccg	2100
ggccccctg	gaaaacctgg	ggctggtgtg	gctggcggtg	gtggccctgg	gggctgtgtg	2160
cctggtgctg	ctgctgctgg	tgtgtcatt	gcgcggcg	ctgcgggaag	agctggagaa	2220

**THIS PAGE BLANK (USPTO)**

```

aggggccaag gctactgaga ggaccttgggt gtacccccctg gagctgcca aggagccac 2280
cagtcacccc ttccggccct gtccctgaacc agatgagaaa ctttgggata ctgtcgggta 2340
ctactattca gatggctccc ttaagatagt acctgggcat gcccgggtgcc agcccgggtg 2400
ggggcccccct tcgccacctc caggcatccc agggcagcct ctgccttctc caactcggct 2460
tcacctgggg ggtgggcgga actcaaatgc caatggttac gtgcgcttac aactaggagg 2520
ggaggaccgg ggagggtctcg ggcacccccct gcctgagctc gcggatgaac tgagacgcaa 2580
actgcagcaa cgccagccac tgcccagctc caaccccag gagtcatacag tatgaggggga 2640
acccccaccg cgtcggcggg aagcgtggga ggtgtagctc ctacttttgc acaggcacca 2700
gctacctcag ggacatggca cgggcacctg ctctgtctgg gacagatact gccagcacc 2760
caccgggcca tgaggacctg ctctgtctcag cacgggcact gccacttgggt gtggctcacc 2820
agggcaccag cctcgcagaa ggcattcttc tctctctgt gaatacaga cacgcgggac 2880
ccccccaccg aaaacttttc aaggcagaag ttccaagatg tgtgtttgtc tgtatttgc 2940
catgtgtttg tgtgtgtgtg tatgtgtgtg tgcacgcggtg tgcgcgcttg tggcatagcc 3000
ttcctgtttc tgtcaagtct tcccttggtc tgggtcctcc tgggtgagtc ttggagctat 3060
gaaggggaag gggctgtatc actttgtctc tccctacccc actgccccga gtgtcgggca 3120
gcgatgtaca tatggagggt ggggtggacg ccttcaaaa ggaagtgtca 3180
ggccttgggg tgggcctagt cttgtctcct agggctgtga atggttttca gagggtgggg 3240
gaaggaaatt ggaacccctc tgtgtgtttt gggggggaaa ggggttggtg ggg 3293

```

&lt;210&gt; 26

&lt;211&gt; 1324

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1689337CB1

&lt;400&gt; 26

```

cagaatatca gatgtccaac atttcccttc tggctctttt tagctcccaa gtatctgctt 60
ccttaaaggc cctgagtcac ttcttctctc tttgcttcag actggccccg gagcaagcgc 120
gagtgtgcga actgcagagt gggaaccagc agctggagga gcagcgggtg gagctggtgg 180
aaagactgca ggccatgctg caggcccaact gggatgaggc caaccagctg ctacagacca 240
ctctcccgcc gcccaacctc ccagctcttc ctgctggacc ctccagcccc gggcctcagg 300
agccccagaa ggaggagagg aggggtctgga ctatgcctcc catggccgtg gccctgaagc 360
ctgtatttga gcagagccgg gaagcaaggg acgagctacc tggagcgct cctgttcttt 420
gcagtctctc ctcatatctt agcctcctgt tgggccccctc ttttcagagc cagcattctt 480
tccagccccct ggagcccaaa ccagacctca ctccatccac agctgggggc ttctctgcac 540
ttggggccct ccateccgat catagggcag aaaggccatt ccctgaggaa gatcctggac 600
ctgacgggga gggcctccta aagcaagggc tgccgcctgc tcagctggag ggcccaaga 660
atttttttga ccagttgctg gagacagtgc ccagaacaa tgagaacctc tctgtcgacc 720
tgtttgcccc taagtctggt cctctgactg tcccatcttg ggaggaagcc cctcaagtgc 780
cacgtatttc accgcctgtc cacaacaaac aagttccctt agccatggca tccagtcttt 840
tccgggtccc tgagcctccc tccctccatt cacaaggcag tgggtcccag agtgggtccc 900
cagagagagg tggagatggg cttacattcc caaggcagct gatggagggtg tctcaactgt 960
tgcgactcta ccaggtctcg ggcctggggg ctctgcctgc tgaggatctc ctgctctacc 1020
tgaagaggct ggaacacagc ggcaggactg atggccgagg ggataatgtc ccagaagga 1080
acacagactc ccgcttgggt gagatcccc ggaaagagat tccctcccag gctgtccctc 1140
gccgccttgc tacagccccc aagactgaaa aacctccgc acggaagaaa agtgggcacc 1200
ctgccccgag tagcatgagg agccgggggg gagtctggag atgagcccc ctacctctc 1260
tctcttttgt tctctcattg ttgttatttt aataaatgtc cagtagtctg taaaaaaaa 1320
aaaa

```

&lt;210&gt; 27

&lt;211&gt; 1884

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1746392CB1

&lt;400&gt; 27

```

cgagggggtg tggaaactta ccggctgagc catggatata ccgttaaggc gcagccgacg 60
gctgggaggc ctaaggcccc aatccccga gagctcacc tcagtttcgc ggacgagacg 120
ggccttctgt gagttcgagt cgaaccaga agaaacgagg gagccgggt ctctccgag 180
tgtgcagcgg gctggccttg ggtccccga aaggccgccc aagacaagcc caggatcacc 240

```

**THIS PAGE BLANK (USPTO)**



```

ccgtctgcag caggggtgcag gcttggagtc accccaaggg cagccagagc cagggcgagc 300
gtccccccag cgtcagcaag acctacacct ggagtcgcct caaagacagc cagagtacag 360
tctgaatcc ccacgatgtc agccgaagcc aagtgaggag gcaccgaagt gttctcagga 420
ccagggagta ctggcctcgg agttggccca gaataaggag gagctgacct cggggggccc 480
ccagcatcag ctaccgccgg tcccaggatc accagagcct taccgccgtc agcaagctcc 540
cggtcaggag cctctcagc cactactgga gctgacaccc agggcacctg gctccccccg 600
gggtcagcat gagccgagca agccacctcc agctggggag acggtgacag gcggcttcgg 660
ggcaaaaga cgaaaagggt tctcatccca ggcccagcg tccaagaagt tgaataaaga 720
ggagcttctt gtaatcccga aggggaagcc caaatcgggg cgagtgtgga aggaccgtc 780
caagaaaaga ttctcccaga tgcttcagga caagcccctg cgcacatcgt ggcagcggaa 840
gatgaaggaa cgacaggaga ggaagctggc caaggacttt gcccgtcacc tggaggagga 900
gaagtgaaag cgccgccagg agaagaaaca gcgcccgggt gagaacctga aacgccgcct 960
ggagaatgag cggaaggcag aggtcgtcca agtgatccga aaccccgcga agctcaagcg 1020
ggcaaaagaag aagcagctgc gctccattga gaagcgggac accctggccc tgctgcagaa 1080
gcagccgccc cagcagccgg cagccaagat ctgagctcag gacggcccga gggccttcca 1140
tggccaaga acatgtcaga cacagcacct cagggcgctg ctcatatgcc tctgtcggag 1200
ctggcactcc aaacccatgg ctccagaaca gggaccccca ccccgaccgg ggctcctcgg 1260
cctttgaagg ctccaggca ggtctgtgtg ggacagaagc ccagaggggg cctgggacct 1320
ggcagagatg ggggcgggaa gagattcagc tcccatccct ccttctctc cttctccaag 1380
tgcttcaaa acaagaactg tacattcttc tggttctcca gtgagctggg gactggcagg 1440
tgactccctc agcagtgtat gccctttctc agcatcctag gtccatccca ggctggagg 1500
ctgacagttg ggaatccagc tccccccaca ccttcccaaa ggctgctctg agcacctcca 1560
caccctactg cctctgtccc cagcaaaactg aatccggttc ctctccactt ttcaataactg 1620
aaagattaaa atggggagggt tgcagggagc agagcttttc cctagcacc actttccaa 1680
accagtctct gcagaagccc cagagaatct aactcatgcc tgtccagttc acagcaaaaa 1740
tatttattga gtgcctgttg catacaggca caatcctagg caccggcaaa tacagacaat 1800
agaccaaagt cctgcctc gaggagcttt cattctgatg gagagaaaac ataataaaca 1860
agcaaaatgc acaaaaaaaa aaaa 1884

```

&lt;210&gt; 28

&lt;211&gt; 2429

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1825182CB1

&lt;400&gt; 28

```

cctatttcca aaaagctcga ctggagtggt ataaaacctg aaaattctct tgtgctttct 60
cttcttttgc ttctagttac catcctcaaa ggattggcta aaagcaagca actggattga 120
acaccctaag aagaaagatt cacactgcac caggagacat cagaaagaat gaaaactctg 180
cogctgtttg tgtgcatctg tgcactgagt gcttgcttct cgttcagtga aggtcgagaa 240
agggatcatg aactacgtca cagaaggcat catcaccaat caccaaaatc tcaatttgaa 300
ttaccacatt atcctggact gctagctcac cagaagccgt tcattagaaa gtccataaaa 360
tgtctgcaca aacgctgtag gcctaagctt ccaccttcac ctaataagcc ccccaaatc 420
ccaaatcctc accagccacc taaacatcca gataaaaata gcagtgtggg caaccctacc 480
ttagtggcta caaccaaaat tccatctgtg actttccat cagcttccac caaaattact 540
acccttccaa atgtgacttt tcttccccag aatgccacca ccatactctc aagagaaaaat 600
gttaacacaa gctcttctgt agctacatta gcaccagtga attccccagc tccacaagac 660
accacagctg cccacccac accttctgca actacaccag ctccaccatc ttctcagct 720
ccaccagaga cccacagctg cccacccaca ccttctgcaa ctacacaagc tccaccatct 780
tctcagctc caccagagac cacagctgcc ccacccacac ctctgcaac tacaccagct 840
ccaccatctt cctcagctcc accagagacc acagctgccc caccacacc ttctgcaact 900
acaccagctc cactatcttc ctgagctcca ccagagacca cagctgtccc acccacact 960
tctgcaacta ccttagaccc atcatccgct tcagctccac cagagaccac agctgcccc 1020
ccacacactt ctgcaactac accagctcca ccgtcttccc cagctccaca agagaccaca 1080
gctgccccaa ttaccacacc taattcttcc ccaactactc ttgcacctga cacttctgaa 1140
acttcagctg caccacaca ccagactact acttcggtea ctactcaaac tactactact 1200
aaacaaccaa cttcagctcc tggccaaaat aaaatttctc gatttctttt atattgaag 1260
aatctactaa acagaattat tgacgacatg gtggagcaat agtatattgt atgttgtaaa 1320
gtgttctgtc atttacaaga tgtgattcat gagtgcagaa ctaccacctt tcttttagca 1380
ccaatcccaa catgaaatta tattactcag atttaaagca ctatcattaa tctttcaatc 1440
taattattca ccaccacaag acctattaac aagacaaaat gcctctatcc cacaagcag 1500
atgcaggtct ggggttcaaa ataactcttt ggatcctaca gagatagcct actgagggca 1560
gagaaagtcc ttagataaag agagaatatt gtatgggcca tcaaccattt acttttccct 1620
gaatgttaga aactacaaaa ccactacctt gtacccccat caaaatccca cctgaacct 1680

```

**THIS PAGE BLANK (USPTO)**

ctaatectat	aaacataaag	gggtaaaatt	ggaactctcc	agatgaacaa	agacatctaa	1740
atatctgtag	atagaaacat	ttatctatct	aaatatattg	atagacctgt	cattgtattg	1800
attaatgaca	aaacccttta	gataattatc	ttccatttta	aataaaaattt	tatttcacaa	1860
atatgagcca	agaaagagga	aagttgattt	gaagtgaagga	ttagaagtga	atgacaataa	1920
agtctggcag	ccaagcacga	accaagactg	gcactatttt	tcttagtgta	tataattgtt	1980
taaactgcaa	ggttgacatt	tattgtgttg	tgtctaagtt	aattttgatc	taatgtacct	2040
gattctagcc	tctgtgaaca	acaagaatat	gtttgtgtat	gttcacatgg	tgcttataat	2100
atttcactat	caattcaatt	aattcacata	aattccatgt	gaaatgtatt	caacaatgga	2160
atattttcta	aaacattttag	tatacatttg	aatgtatttt	aaacctgccc	aaactactgc	2220
tttaatgtca	agtttgcaga	attgtctctg	aaaataaaaa	ccctgacttt	agttgtaaaa	2280
caataaaagt	tagtacttg	gtatacggag	atgttaattt	gggatattgga	ggcattttta	2340
tcttctgtca	ctactactta	aaactctgat	gattatgtta	gatttttttg	ctaactaata	2400
aagatttcaa	atggcaattt	ataaaaaaa				2429

<210> 29  
 <211> 985  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2155541CB1

<400> 29						
ggaagagggc	ggcgacgggtg	gtgggtgactg	agcggagccc	ggtgacagga	tgttgggtgtt	60
ggtattagga	gatctgcaca	tcccacaccg	gtgcaacagt	ttgccagcta	aattcaaaaa	120
actcctgggtg	ccaggaaaaa	ttcagcacat	tctctgcaca	ggaaaccttt	gcacaaaaga	180
gagttatgac	tatctcaaga	ctctggctgg	tgatgttcat	attgtgagag	gagacttcga	240
tgagaatctg	aattatccag	aacagaaagt	tgtgactgtt	ggacagttca	aaattgggtct	300
gatccatgga	catcaagtta	ttccatgggg	agatatggcc	agcttagccc	tgttgcagag	360
gcaatttgat	gtggacattc	ttatctcggg	acacacacac	aaatttgaag	catttgagca	420
tgaaaaataa	ttctacatta	atccagggttc	tgccactggg	gcatataatg	ccttggaac	480
aaacattatt	ccatcatttg	tgttgatgga	tatccaggct	tctacagtgg	tcacctatgt	540
gtatcagcta	attggagatg	atgtgaaagt	agaacgaatc	gaatacaaaa	aaccttaag	600
ccaggcctgt	cttgatgatt	tttggttttt	tttcattgtc	ctgttgaaat	caagtaatta	660
aacatttaag	agccacaaaa	ttgtatcact	tttataatat	tttgagtaaa	aatataatc	720
catcttctct	gttaatacat	aattgctcca	agcttctgt	aaactataag	aatatattta	780
gtttacagta	tatggattct	atgaaaaaat	gtccacaaca	cagtaattgg	tcacttgta	840
agaaaaattt	atccttgtaa	gtatcttcaa	agttgatatt	tggaacttta	ttccaaaagt	900
agtgcatgtg	gagaaagaat	ctagactttc	ttgtatacat	ttttctcttc	tccagtaata	960
aacaattacc	tttcattta	aaaaa				985

<210> 30  
 <211> 3381  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2215706CB1

<400> 30						
gttagctcag	tgtaacattg	atgagctgaa	gaaagagatg	aatatgaatt	ttggagactg	60
gcaccttttc	agaagcacag	tactagaaat	gagaaacgca	gaaagccacg	tggtccctga	120
agacccacgt	ttcctcagtg	agagcagcag	tgcccagcc	ccgcacgggtg	agcctgctcg	180
ccgcgcttcc	cacaacgagc	tgctcacac	cgagctctcc	agccagacgc	cctacacact	240
caacttcagc	ttcgaagagc	tgaacacgct	tggcctggat	gaagggtgcc	ctcgtcacag	300
taatctaagt	tggcagtcac	aaactcgcag	aaccccaagt	ctttcgagtc	tcaattccca	360
ggattccagt	attgaaattt	caaagcttac	tgataagggtg	caggccgagt	atagagatgc	420
ctatagagaa	tacattgctc	agatgtccca	gttagaaggg	ggccccgggt	ctacaacat	480
tagtggcaga	tcttctccac	atagcacata	ttacatgggt	cagagtcat	cagggggctc	540
tattcattca	aacctagagc	aagaaaaggg	gaaggatagt	gaaccaaagc	ccgatgatgg	600
gaggaagtc	tttctaata	agaggggaga	tgttatcgat	tattcatcat	caggggtttc	660
caccaacgat	gcttcccccc	tggatcctat	cactgaagaa	gatgaaaaat	cagatcagtc	720
aggcagtaag	cttctccag	gcaagaaatc	ttccgaaagg	tcaagcctct	tccagacaga	780
tttgaagctt	aagggaagtg	ggctgcgcta	tcaaaaactc	ccaagtgcag	aggatgaatc	840
tggcacagaa	gaatcagata	acactccact	gctcaaagat	gacaaagaca	gaaaagccga	900

**THIS PAGE BLANK (USPTO)**

agggaaagta	gagagagtgc	cgaagtctcc	agaacacagt	gctgagccga	tcagaacctt	960
cattaaagcc	aaagagtatt	tatcggatgc	gctccttgac	aaaaaggatt	catcggattc	1020
aggagtgaga	tccagtgaag	gttctcccaa	tcactctctg	cacaatgaag	tggcggatga	1080
ctcccatctt	gaaaaggcaa	atctcataga	gctggaagat	gacagtcaca	gcggaagcgc	1140
gggaatccca	catagcctga	gtggcctgca	agatccaatt	atagctcgga	tgtccatttg	1200
ttcagaagac	aagaaaagcc	cttccgaatg	cagcttgata	gccagcagcc	ctgaagaaaa	1260
ctggcctgca	tgccagaaag	cctacaacct	gaaccgaact	cccagcaccg	tgactctgaa	1320
caacaatagt	gctccagcca	acagagccaa	tcaaaatttc	gatgagatgg	aggggaattag	1380
ggagacttct	caagtcatct	tgaggcctag	ttccagtccc	aacccaacca	ctatttcagaa	1440
tgagaatcta	aaaagcatga	cacataagcg	aagccaacgt	tcaagttaca	caaggctctc	1500
caaagatcct	ccggagctcc	atgcagcagc	ctcttctgag	agcacaggct	ttggagaaga	1560
aagataaagc	attctttgag	aaaaacaagc	aaaggagaag	agtgttactg	taccttatgc	1620
acagaattgt	cctggatttt	gactccatcc	acgcccatac	cctttctaca	ttttgctgac	1680
agataactaa	ccgatgatga	ggccgaggta	aaagagacat	ctgcagtgtg	acagaaggga	1740
gcatgagaag	catggctcac	cagccagcct	ctgtggctct	tgttaattaga	agcttcagaa	1800
ctcacaatac	tttcattgac	tttcattgac	gtcattcccc	ataaaacttt	ttgaccagag	1860
gtgagatctg	agtataaaga	taggtcagaa	gtatttttaa	gggcttaaat	tgccaaaaag	1920
aaaaaaaagc	agagaccctt	tttgcaaaac	tttggtgacc	acacatttga	gggaagacgt	1980
ggcgttaggt	gaagcagaag	caaaccctgc	tcttaggggc	tcacctaggt	gagtgcacag	2040
cctgtgagcg	tacagggaga	ggctgagtaa	accgagatcc	agcgttctgt	atggcagggg	2100
tattgcttat	cacagagggt	ctgaagagta	ggaagtacat	aatgaagagg	gctttaaaaa	2160
ttgccaacaa	agtgaagtcac	cagggctggc	agtagtgtga	cggggctgtc	ctgagctggt	2220
aggagagtag	atgcggggag	ggctgggtgac	ctccgtgggt	ttatatgtcg	gaaactcttc	2280
tctccaaatc	ccaggcctgg	cttcagcac	catccagctg	tgcccaagaa	gccaccctgg	2340
tctgttctcc	aactctttta	aatgggtgcc	aacttttcta	agtgaagctt	gcaatgagaa	2400
gaaaaaaaaa	catgaattct	ttttctggaa	aatcagggag	acatgggtaa	taataggtag	2460
taataaataa	ttatagatga	gtgaatgagg	aaataattac	atcaaaaagg	tcagtgcacaa	2520
ttgataaatg	acaaggaaat	atttaattag	gtaaaactaa	atcattgtct	tctatactag	2580
gatagacttt	atctacttca	tctgttctca	agtcagcatg	ttagttctgg	ggaaggatca	2640
taagaaagga	aatacttttt	aaaaaaaaat	ttggaaacat	gtaacaaagc	aagggtaaaa	2700
tatatatata	tatctatata	agtgtctgtg	ctgtaaaagt	gtactttcca	ttaattatta	2760
gccgagttaa	gagaatggtc	acattgaagt	actgtgtgga	ctagaaatgt	accctgtcat	2820
catgcaatga	aatattgtta	tcgttttaac	atagctcatt	tatgtagaat	gaattctggg	2880
gggtttacccc	aagtcacagt	taggacggta	gatggtgaga	tcgcagatgc	gctattatct	2940
agattcagtg	ttacattttc	gatgtttatc	actcagtggg	tttttattaa	tatgctgatt	3000
aagttattta	ctgggcccag	cattgtgcta	aatagttgct	cttttgtgtt	tcattgcctt	3060
gatgtttgag	tgtaatctag	catttttaata	cagtgtttat	tttgcattat	ctttaacaaa	3120
tgttttaagc	aatttttaaaa	aggcaggatg	ttattgacat	tataactga	agtcttaaca	3180
ttttaacatt	tatagtgcct	atttgcaaaa	ttgtataatt	aggaattatt	tcagagacaa	3240
tgttttcttt	ttcagggtgag	tagttgccgc	gtaatatcat	tggagtacat	tctttatact	3300
gtttgtgaaa	tttaactatg	catattaagt	gtacaaatag	atttagaaaa	caataaaaaa	3360
ttgcatgcta	aaaaaaaaaa	a				3381

<210> 31  
 <211> 1803  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2347692CB1

<400> 31	gggctgttcc	cggggaggct	gtgatgggtt	gacagggtgcg	tgacagtggg	agctgctctc	60
	ggcacaagca	tgtacggcaa	aggcaagagt	aacagcagcg	ccgtcccgtc	cgacagccag	120
	gcccgggaga	agtttagcact	ctacgtatat	gaatatctgc	tccatgtagg	agctcagaaa	180
	tcagctcaaa	cattttttatc	agagataaga	tgggaaaaaa	acatcacatt	gggggaacca	240
	ccaggattct	tacattcttg	gtgggtgtgta	ttttgggata	tctactgtgc	agctccagag	300
	agacgtgaaa	catgtgaaca	ctcaagtga	gcaaaaagcct	tccatgatta	cagtgtctga	360
	gcagctccca	gtccagtgct	aggaaacatt	ccccaggag	atggcatgcc	agtaggtcct	420
	gtaccaccag	ggttctttca	gccttttatg	tcacctcggt	accctggagg	tccaaggccc	480
	ccattgagga	tacctaatca	ggcacttgga	gggtgtccag	gaagtcagcc	attactcccc	540
	agtggaaatg	atccaactcg	acaacaagga	catccaaata	tgggtgggccc	aatgcagaga	600
	atgactctc	caagaggaat	gggtgccctta	ggaccacaga	actatggagg	tgcaatgaga	660
	ccccactga	atgcttttagg	tggccctgga	atgcctggaa	tgaacatggg	tccagggtgg	720
	ggtagacctt	ggccaaaccc	aacaaatgcc	aattcaatac	catactctc	agcatctcct	780
	gggaattatg	taggtcctcc	aggaggtgga	gggccaccag	gaacacccat	catgcctagt	840

**THIS PAGE BLANK (USPTO)**

ccagcagatt	caaccaactc	tggtgataac	atgtatactt	taatgaatgc	agtacctcct	900
ggacctaaaca	gacctaat	tccaatgggt	cctggggtcag	atgggtcccat	gggtggatta	960
ggaggaatgg	agtcacatca	catgaatggc	tctttaggct	caggagatat	ggacagtatt	1020
tccaagaatt	ctcccaataa	tatgagcctg	agtaatacaac	cgggcactcc	aagggatgat	1080
ggcgaaatgg	ggggaaat	cttaaactcct	tttcagagt	agagttactc	ccctagcatg	1140
acaatgagcg	tgtgatccat	taccaagtct	cctcatgaaa	accacagtga	gtcagccctt	1200
cacagaacta	ctacggaaga	aaattattca	tcacagtgt	cagttaaaca	aaggaaatctc	1260
agtcacacca	aaccaac	tttatttctc	gctctctccc	ctcttttgtg	aagacagcgg	1320
gtccaaatgt	gattcaaa	actgtacgga	gtggcatatt	agaattgccc	taaactgaac	1380
tgcaataat	tatgtgtgt	tgtatatgtg	tgggaaagag	aatgtactgt	atatgtgtat	1440
gttatacaga	catatacaca	tacatacat	gacccacagg	acattgtaaa	atattatcac	1500
atgacatctt	aagtagaaat	aagtaggac	ttttattcca	tcctttttt	cacgtttaca	1560
ttttaattat	tacaagttgc	tcctgcccc	tcctgaact	attttgtgct	gtgtatatca	1620
ctgctttata	taagttat	tttaagggtga	actcagatgt	tatgggtttt	taaatgtctg	1680
caatcatgga	taggaataaa	atcgcttatt	tgagagcttt	cattaaattg	cgtctgatgc	1740
aagtatcct	gtgaatcaca	aagtgtactg	tctcaaaaaa	tagaagaaaa	tgtgaaaaaa	1800
aaa						1803

&lt;210&gt; 32

&lt;211&gt; 1515

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2579048CB1

&lt;400&gt; 32

ggtaaagtgg	cttctggg	gaaggtacac	tataggctcg	gggaggtaag	cggcggcagg	60
cggcggttg	gtgtgtccc	ggtgtgggga	ggcgacagag	ccctggcact	tgagggttga	120
gggggcctcc	ccagcgcggc	gaaccgtcta	gcctccggag	gccaggccgt	gagtgcggga	180
gggtatagcc	aaggcggaag	aatttttgcca	ctcactacct	gtgtgacctc	gggtaaaatta	240
gccttggaa	gtcagtttct	tcgtctctat	aattgaaata	ataatagtag	ctctctcagg	300
attgtttgtg	gccgtcagtg	aaacacttag	agcagtttct	ggcacatggt	agaattgggc	360
tatttgcctg	agcttcttgg	tggcccttgc	tagcccagga	agaaacttac	attttgattt	420
ttttgtacca	tggctttgg	tcacaaaattg	ctgcatggta	cttattttct	cagaaaaattc	480
tctaagccaa	cttctgcctt	gtatccattt	ttgggtattc	tctttgcaga	gtattccagt	540
agtcttcaga	aaccagtggc	ttctcctggc	aaagcctcct	cacagaggaa	gactgaagg	600
gatttgcga	gagatcacca	gaaagaagtt	gctttggata	taacttcttc	tgaggagaag	660
cctgatgta	gtttcgataa	agcaattaga	gatgaagcaa	tataccattt	taggcttttg	720
aaggatgaaa	ttgtggatca	ttggagagga	ccggaaggcc	accctctgca	tgagggtctt	780
ctggaacaag	ccaaggttgt	ctggcaattc	cgggggaaag	aagatttgga	taagtggaca	840
gtgacttctg	ataagacgat	tggaggcaga	agtgaagtgt	ttttgaaaat	gggcaagaat	900
aaccaaagt	cactgctata	tggaaactctg	agctctgagg	cgccctcagg	cggggagctc	960
acccgaagt	ggtactgtgc	aatgatatac	aggattccaa	ggggtgcttt	tgagagggaag	1020
atgtcttacg	attggtccca	gttcaatact	ctgtatctcc	gtgtacgtgg	ggatgggtcgg	1080
ccttggatgg	tgaatatcaa	ggaggacaca	gatttcttcc	agaggacgaa	tcagatgtat	1140
agttacttca	tggtcacccg	cggggggacc	tactggcagg	aggccaagat	tcctttttcc	1200
aaatttttct	tctctaatac	aggaagaatc	cgggatgttc	agcatgagct	tcctgttgat	1260
aagatctctt	ctataggatt	caccttggct	gataaagtgg	atgggtccatt	cttctgggag	1320
atagatttta	ttggcgtgtt	tactgatcca	gctcatcacg	aagaatttgc	ctatgaaaaa	1380
tctccagag	ttaaaccaag	gctttttaaa	taaagatcat	atggtagttt	tgttttacta	1440
atctaagggt	actagcatct	acaatgatata	agacaaaata	aaatatattt	ttaatggcat	1500
caaaaaaaaa	aaaaa					1515

&lt;210&gt; 33

&lt;211&gt; 4416

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2604493CB1

&lt;400&gt; 33

cggatggg	ggcagccg	ccgcccgc	acttttttaa	ttttttcgg	tgccgcagc	60
gcgacccct	ggcgccgat	tccctgatcc	ctggagcgac	gacggccgct	gcctaagctg	120

**THIS PAGE BLANK (USPTO)**



gaaagaggaa	tgccagctcc	tgagcaggcc	tcattggtgg	aggaggggca	accacagacc	180
cgccagggaag	ctgcctccac	tggcccaggc	atggaacccg	agaccacagc	caccactatt	240
ctagcatccg	tgaaggagca	ggagcttcag	tttcagcgac	tcacccgaga	actggaagtg	300
gaaaggcaga	ttgttgccag	tcagctagaa	agatgtaggc	ttggagcaga	atcaccaagc	360
atcgccagca	ccagctcaac	tgagaagtca	tttccttgga	gatcaacaga	cgtgccaaat	420
actggtgtaa	gcaaaccctag	agttttctgac	gctgtccagc	ccaacaacta	tctcatcagg	480
acagagccag	aacaaggaac	cctctattca	ccagaacaga	catctctcca	tgaagttag	540
ggatcattgg	gtaactcaag	aagttcaaca	caaatgaatt	cttattccga	cagtggtatc	600
caggaagcag	ggagtttcca	caaagccag	aacgtgagca	aggcagacaa	cagacagcag	660
cattcattca	taggatcaac	taacaacat	gtggtgagga	attcaagagc	tgaaggacaa	720
acactgggtc	agccatcagt	agccaatcgg	gccatgagaa	gagttagttc	agttccatct	780
agagccacag	gccttcttta	tgttatcagc	acaggcgtgt	ctccttcaag	ggggtctctg	840
agaacttctc	tgggtagtgg	atttggctct	ccgtcagtga	ccgacccccg	acctctgaac	900
cccagtgcat	attcctccac	cacattacct	gctgcacggg	cagcctctcc	gtactcacag	960
agaccgcct	ccccaacagc	tatacggcgg	attgggtcag	tcacctcccg	gcagacctcc	1020
aatcccaagc	gaccaacccc	tcaataccca	accagccca	gagtggggtc	cccactgact	1080
ctgacggatg	cacagactcg	agtagcttcc	ccatcccaag	gccaggtggg	gtcgtcgtcc	1140
cccaaaccgt	cagggatgac	cgccgtacca	cagcatctgg	gaccttcaact	gcaaaggact	1200
gttcatgaca	tggagcaatt	cggacagcag	cagtagaca	tttatgagag	gatgggtcca	1260
cccagccagc	ccctctgac	aggcttacgg	agttctatg	ctagtacaga	tagtgcagct	1320
gggcaagacc	ttcgttctgc	cgtgtctccc	gacttgacac	ttactcctat	atatgagggg	1380
aggacctatt	acagcccagt	gtaccgcagc	ccaaaccatg	gaactgtgga	gctccaagga	1440
tcgcagacgy	cgttgtatcg	cacaggttca	gggtattgga	atctacaaag	gacatccagc	1500
caacgaagac	cccttataca	caaagaaat	aattatgctc	tgaacacaac	agctacgact	1560
gcggagccct	acaggccctat	acaataccga	gtgcaagagt	gcaattataa	caggcttcag	1620
catgcagtgc	cggctgatga	tggcaccaca	agatcccat	caatagacag	cattcagaag	1680
gacccagggy	agtttgccctg	gcgtgatcct	gagttgcctg	aggctcattca	catgcttcag	1740
caccagttcc	catctgttca	ggcaaatgca	gcggcctacc	tgcagcacct	gtgctttggt	1800
gacaacaaag	tgaagtggga	gggtgtgtag	ttagggggaa	tcaagcatct	gggtgacctt	1860
ctggaccaca	gagttttgga	agttcagaag	aatgcttgtg	gtgcccttcg	aaacctcggt	1920
tttggcaagt	ctacagatga	aaataaaata	gcaatgaaga	atggttggtgg	gataacctgcc	1980
ttgttgcgac	tgttgagaaa	atctattgat	gcagaagtaa	gggagcttgt	tacaggagtt	2040
ctttggaatt	tatctcatg	tgatgctgta	aaaatgacaa	tcattcgaga	tgctctctca	2100
accttaacaa	acactgtgat	tgttccacat	tctggatgga	ataactcttc	ttttgatgat	2160
gatcataaaa	ttaaatttca	gacttcacta	gttctgcgta	acacgacagg	ttgcctaagg	2220
aaacttcagc	ccgcggggga	agaagctcgg	aagcaaatgc	ggctctgcga	ggggctggta	2280
gactcactgt	tgtatgtgat	ccacacgtgt	gtgaacacat	ccgattacga	cagcaagacg	2340
gtggagaact	gcgtgtgcac	cctgaggaac	ctgtcctatc	ggctggagct	ggaggtgcc	2400
caggcccggg	tactgggact	gaacgaattg	gatgacttac	taggaaaaga	gtctccagc	2460
aaagactctg	agccaagtgt	ctgggggaag	aagaaagaaa	agaaaaagag	gactccgcaa	2520
gaagatcaat	gggatggagt	tggctctatc	ccaggactgt	cgaagtcccc	caaaggggtt	2580
gagatgctgt	ggcaccatc	gggtgtaaaa	ccatatctga	ctcttctagc	agaaagtccc	2640
aaaccagcca	ccttgggaagg	ctctgcaggg	tctctccaga	acctctctgc	tggcaactgg	2700
aagtttgctc	catatatccg	ggcggccgct	cgaaaagaaa	aggggctccc	catccttggt	2760
gagcttctga	gaatggataa	cgatagagtt	gtttcttccg	tggcaacagc	cttgagggaat	2820
atggcactag	atgttcgcaa	caaggagctc	ataggcaaat	acgccatgcy	agacctgggtc	2880
aaccggctcc	ccggcggcaa	tggccccagt	gtcttgtctg	atgagaccat	ggcagccatc	2940
tgctgtgctc	tgcacgaggt	caccagcaaa	aacatggaga	acgcaaaagc	cctggccgag	3000
tcaggaggca	tagagaagct	gggtgaacata	accaaaggca	ggggcgacag	atcatctctg	3060
aaagtgggtga	aggcagcagc	ccaggctcttg	aatacatat	ggcaatatcg	ggacctccgg	3120
agcatttata	aaaaggatgg	gtggaatcag	aaccatttta	ttacacctgt	gtcgacattg	3180
gagcgagacc	gattcaaatc	acatccttcc	ttgtctacca	ccaaccaaca	gatgtcacc	3240
atcattcagt	cagtccggcag	cacctcttcc	tcaccagcac	tgttaggaat	cagagaccct	3300
cgctctgaat	acgataggac	ccagccacct	atgcagtatt	acaatagcca	aggggatgcc	3360
acacataaag	gcctgtaccc	tggtaagacg	ccagttgggt	gtgtgatata	gtctcttgaa	3420
aagccgcagt	tccaggcgct	tggccagtg	cctgggaagt	agcctgtgct	tgtattgaga	3480
cagtcaccca	gcagcaaac	atgttccagt	cattcccttt	cctactttgg	ggattgttgc	3540
ctttctctgt	tgtttaaagt	aaaacaagca	tgtacttgtt	tgtatgtatg	tatgtatgta	3600
gttgtacggg	gggcacaaat	aaaaagaggg	ctgtatccaa	ataaatcatt	tctggctgct	3660
cactggcaca	gtccccttgc	tccgtcccc	cctggctcga	gcagtctctg	tgtttccac	3720
atgcatacaga	accgtcagcc	cagtgtcact	ttgcaggggc	cacatcttct	agactggctc	3780
atctcttaaa	ttcaaaccag	agaatggaga	tcaatttcat	ttacatttct	atggagaaga	3840
gtttgaataa	atagcaaaag	gtataatgtg	attcttctca	acgtgattat	tttttaagaa	3900
taacaacaaa	aaataatggt	acatagtgtt	aatatttcta	tatcctattc	agaaagctgt	3960
gcatacaatt	tatttaagg	ataccagccc	aaaggggtac	taagatttca	tactgttccc	4020
attctcaacc	gtctggcgat	tataactact	ttagatgttc	caaataagtc	agtgtgagct	4080
accccaaaaa	aagtactctt	ttctcaaaac	tgtctattaa	aaataaataa	aagctggtgg	4140

**THIS PAGE BLANK (USPTO)**

cagcttttact	cctgctgggg	gggcagtcctc	cattgtactg	ttgtctttga	aattgtgatt	4200
tcacaggact	atttactaat	cgactgttgt	tactatggca	attaatagct	cagaaaaattg	4260
catatgctac	atttcaacac	caagctcttg	gcaaaaaggaa	atgcacatac	caactgaagt	4320
gcatgggttg	ttccgatttg	cagcagcctg	ctttatctta	cagatcacct	cttttagcatg	4380
catgcccttc	tgtatcttct	tcgcttgctt	caataa			4416

&lt;210&gt; 34

&lt;211&gt; 4428

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2787182CB1

&lt;400&gt; 34

cggagggcgg	gggctgggtg	gcagcggcca	tggaccgctt	cgtttggacc	agcggcctcc	60
tggagatcaa	cgagaccctg	gtgatccagc	agcgcggggt	gcgaatctac	gatggcgagg	120
agaagataaa	atttgatgct	gggactctcc	ttcttagtac	acaccgactg	atttggagag	180
atcagaaaaa	tcacagtggt	tgcacggcca	ttctccttcc	ccaaattgtg	ttcattgaag	240
aacaggcggc	tggcaattgg	aagagtgcga	aaatagtggg	tcaccttcac	ccagctcctc	300
ctaacaaaga	acctggccca	ttccagagta	gtaagaactc	ctacatcaaa	ctctccttca	360
aagaacatgg	ccagattgag	ttttacaggc	gtttatcaga	ggaaatgaca	caaagaagat	420
gggaagatat	catcagtttc	cagtcattac	aaacaaatag	aggacccag	ccaggaagaa	480
taagggctgt	aggaattgta	ggtattgaaa	ggaaactgga	agaaaaaaa	aaagaaactg	540
acaaaaacat	ttctgaggcc	tttgaagacc	tcagcaaact	aatgatcaag	gctaaggaaa	600
tgggtggaat	atcaaaatca	attgctaata	aaattaaaga	caaacaaggt	gacatcacag	660
aagatgaagac	catcaggttt	aaatcctact	tgctgagcat	gggaatagct	aaccagtta	720
ccagagaaac	ctacggctca	ggcacacagt	accacatgca	gctggccaaa	caactggctg	780
gaatattgca	ggtgccttta	gaggaacgag	ggggaataat	gtcactcacg	gaggtgtact	840
gcttagtaaa	ccgagctcga	ggaatggaat	tgctctcacc	agaagattta	gtgaatgcgt	900
gcaagatgct	ggaagcactg	aaattacctc	tcaggctccg	tgtgtttgac	agtggcgctca	960
tggtaattga	gcttcagctc	cacaaggaag	aggaaatggg	ggcctcggcc	ctggagacag	1020
tttcagaaaa	gggatcccta	acatcagaag	agtttgctaa	gcttgtggga	atgtctgtcc	1080
tcctagccaa	agaaagggtg	ctgcttgag	agaagatggg	ccatctttgc	cggtgtgtac	1140
cagtgaagag	cctgcgtttt	tacccaaatt	tatttatgac	acagagctaa	gggttttgta	1200
tttaaaatcc	tttttgtcca	tatgcttgcg	tcagttagag	gttgtagtac	attgagctaa	1260
gagataaacc	ccgatcaatt	gagaatttat	tagaacttca	cagtgcattg	taaatctctt	1320
ttaatttctc	actaaatatg	gtccaggaaa	tttatttagt	atacgcatag	gaaaatctag	1380
aaaagtgaat	cccaatatga	atttaaaatc	atgctatagt	gcagaaccct	cagagttaa	1440
cttggaaat	agtggatttt	aacttgatcc	tcaaactctaa	tcattttata	aagaaggga	1500
tttagttttg	cagagaataa	aaagagaagt	tgcattgtca	gacagggttag	attattattt	1560
tgggtgtaact	gaaattcact	gattgcacat	gacaatgttg	ggacaaaata	tactcagca	1620
tgctatatga	ggctcctccc	cagggctttt	agaagcagtc	atagacatgt	cttcaacata	1680
ccaaataaaa	tacctttaaa	aatgaaataa	ttttatttga	cacatatatt	tatatattt	1740
ctatctagtt	tctctttgtt	tttttaagt	atgatttcat	gactggcatt	taaagaatgc	1800
aactgtgtca	tttgttttcc	aaatgctgtg	gattttgaaa	ctgaactaga	gagctgtata	1860
gacatgccca	gagtattgat	tacaaattta	ggaggtagac	ggctcaggaa	ttccctggga	1920
ttgttgtgct	ggtggaatgg	cagagggaac	ttcacaggaa	ccttagtgtt	cttttacctc	1980
aaagccacag	acaggaaata	gaaagtggaa	aagtaatatc	tccttttctt	ttccataagg	2040
agtttcaaca	ctgaacttta	aaaagtctat	catattccag	caatattttt	tctttgtcct	2100
ttatgttgta	agttgtgtgg	aaaaactact	tcggtaagaa	atgttactga	gataacaaca	2160
actggctaata	actgcatgta	gattgcttag	gttttaaaagt	gactgcctga	cttcacatgt	2220
tattgtctaca	gcctccagta	tgttcgcatt	atctcaaact	cagggacccc	acaggacagg	2280
agacaccctt	tctgaaaactg	agttggaagt	gaaagggttg	tgatggtttt	ggccaagcct	2340
gcgggagggga	aagtattgta	ttgggagcac	ccttgggacc	aggaagaggg	atgccaggt	2400
tcacactctg	ggacccctaa	gacattgtca	gtgggtaagg	tggaggggcca	ctgccagagg	2460
ctgctgcagt	gccctctggg	aagtgccagg	gacctcagc	tgacagacca	ggggatccca	2520
atcccttatc	tctgcttgaa	aagtggctat	tgctgcttcg	tatcctccag	tgccacagga	2580
atgctgtgcc	tgagacatcc	cacttttagtt	tttgttttcc	aaacctgtca	ctgactcttc	2640
attgtgggtga	cagatcctct	ggttccccca	ttgctcagcc	tcaggtttat	tttgagggtta	2700
actcagtgtc	tgactgcccc	agtcttttag	gggttagttg	gaccatttgtg	taagcttgtt	2760
tgtatgtctc	acctctttct	ggttggtact	tactgtctca	cactgtctctg	gcaaacttgt	2820
acacacacac	acactctctc	tctctctcac	ctggctaaag	cttttaaaat	tatgaagata	2880
aataatctgt	ttcaccagct	ggagtgaagt	gcaaggaaga	ttgctggagc	tactcaatct	2940
agcgctaattg	gtttggattc	attactgcaa	acctacataa	tttaacatat	ttgtttactt	3000
tagtgtgaca	actgatgaaa	aaaaatggag	caatctgaat	tgtataaaat	aacttaagaa	3060

**THIS PAGE BLANK (USPTO)**

ggaagaaaag	tgatatataa	atatatTTTT	caaagtgcac	attaatttaa	aaatgagtat	3120
gattgatttt	atTTTTaaag	tgggcattct	tcactgtttc	gagacctttg	tatgtattttg	3180
tgatTTTTta	tctTTTTttt	tcaggccatt	attataaggt	gttatttttg	ccctctaagt	3240
tagaagttat	gtttaaatac	aaccaatcag	gccctaagt	caaaaaagtt	ctgtcttggt	3300
gcttatcaact	ggtataatta	ttatTTTTtt	gttttccctaa	ctttgctctt	aggagcatga	3360
gcctctttgt	agctttatgg	taatgcaata	tttctggtca	tttactgtca	aaaaattttt	3420
ttacattgtg	ttagtgaaga	ctgtgttttc	agggttaaata	gattgggtatc	aatgtatata	3480
gagtaaaatt	attgcaaaat	ttaaaaaatga	tttttcttgg	cattcttttt	aaatactctg	3540
aaacatatat	gcaaatgaga	aacctttaat	catactaaag	ccagttatgt	taagaatttt	3600
tcctttggat	tttataatta	agagtgtgct	ataaatactt	atgacattga	gctcttactg	3660
ttttagtgtg	cttaaaaagaa	tgggacatgt	catgaactgg	tttctttttt	taaaaaattc	3720
ttgatgtgaa	ttccactgtt	acctgaccat	gtatattcag	atgaatgtac	tctttcttgt	3780
tccttgttac	cagattaatc	tgagagaatg	tggattcatc	ctgtcatctt	tctcactctc	3840
aaatacttaa	aatggttctt	ttgccccaaa	gtctcagagt	cttcattttg	gtacagtggg	3900
gtcacaaatt	gtaagggtgt	taggttttaa	agaccgtgcc	taaggacaaa	attgagcatg	3960
ctcagaatgg	tccttgaaaa	tctcctaaat	taccaataaa	tttctctttg	tagtctttaa	4020
gtttaaatcc	ctgtggagtg	atatgtaggt	atgaatatat	tatgcttata	ttaagtacaa	4080
aaatacaaac	tgcatatcaa	gagattctta	tagcattaat	aatttccatg	catgtgtctt	4140
tttcagtag	gtatggttga	atltatgtat	atttattgct	aatcccatcc	cttaccgatc	4200
agagtataag	ctgcgcaagg	gcagaagttt	ttatctgggt	tggtcatgga	tgtattctaa	4260
gagctgagaa	cagggcctgg	acacaataag	cattcaataa	atatttactg	aatgaatgaa	4320
ctcctaccta	tattcctatt	tataatttgg	ctccacttta	tcctacttta	gctcccatcc	4380
aattcaataa	aaaaaacatt	tttttgaaca	cataaaaaaa	aaaaaaa		4428

&lt;210&gt; 35

&lt;211&gt; 1907

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3096668CB1

&lt;400&gt; 35

gcgttgctgc	gtaaatggcg	ggggcgtgtc	ttttggctcc	tcgcggtgta	gttacctgag	60
aaacgcggga	agttgggccc	aggcagtggt	gctgcgggtg	cctaagttgt	tttactattt	120
ctggagagag	ccgtgagctt	gtccaggggc	cccaatcctg	aggccgaccc	ggtttctggc	180
gcggtgcatg	ggaggtagtg	gagggccgcg	ccgctcagct	ggaaactctg	aaattcaatg	240
gcaccgactt	tggagtggg	gaaggctcgg	cggctccgct	tcggggctct	gcccctgtgc	300
caggggacaca	gccgcccgtc	cagtcgtttg	aggggtcccc	ggacgctggg	cagaccgtgg	360
aggttaagcc	tgccggggag	cagcctctgc	agcccgttct	gaacgccgtc	gcggccggga	420
ccccggcgcc	gcagccacag	ccaccggctg	aatcgccggc	ctgcgggagc	tgcgctcacct	480
ccccaggagc	cgcagagcct	gcgcgggcgc	cggactcctt	ggagacctcg	gactcggatt	540
cggactcgga	cagtgaaaca	gattcagata	gttcaagttc	atcgctcttc	tcttcattct	600
cctcatcgtc	gtcttcttcc	tcttgatat	cacttctctc	agtgtgtgta	gatggagatg	660
atgatttaca	agttgagaag	gaaaaataaga	attttctctt	taaaaacaaa	gatgaattac	720
ttcttaactc	atagagaaaa	aaaaaaattca	gtcatggctc	tcagtactcc	agggatttga	780
tcctactttc	caaccatata	tgaaaagtga	aggctcacct	aatagcatca	ggtaattggc	840
tgccacgtgc	atgtcctgtg	ttgggtaata	tgagaaaagt	ggtagataaa	tttcatccca	900
ccttctctac	ttttctcaat	tctcccaaaa	ctgacaccaa	aaagggttag	ttttcatttg	960
gaatgtttga	ttccttctcc	caccaacctc	tggtgtcctt	tttccatatt	tgtgagtga	1020
aatcctactg	aaagtctctt	atataatttt	acttttagaaa	gtcatcaatt	tactgatact	1080
cttttttgct	caataagcct	tcacaaatgc	ttgaaaagtc	aacagtgtgg	taaatgataa	1140
atgttacctt	ttcagagggc	cacttttgac	atattagctt	cttccatttt	tgcatgtgaa	1200
gtgttaggca	cccatTTTT	gactggtagt	tctataacct	cattttttaa	taggatggat	1260
tttttaaatg	tatgtggtac	tcttttgaat	attttgctta	actgaataac	ctttagtagc	1320
atacagatgt	tagaaaaatta	aaatataata	atcttacaaa	aacagacaat	aattgctctt	1380
tcattggatat	tcagaaagt	tggttaggat	cttgcatctc	tttaatgcca	tcgttacgag	1440
tatctgttat	cacagaagca	cagtaatttt	tttattggga	tctagtatat	aatttctgta	1500
ttaatgtgtt	gtctttttga	ttacatgcag	ttggacattt	aaattgtcgt	tagccatttg	1560
tgtaactgtt	tagaacttgt	gtcagtttc	cttactattc	gtaggtaaga	aaaatgaaac	1620
cagaaagaaa	gagacttgct	tgaagtcata	ttagtgggga	ggagcagatc	tgaactggta	1680
atagatacat	tccaaactgc	tgtgttttta	gtatttgggt	agacctgcat	agcaacgttg	1740
gaatctggtc	gttttgactt	tctcattatt	tggtttctga	ataatcatgt	agcaattttt	1800
ctcaaagtga	aggatgaatg	tatcatctaa	tgtaagcttt	ttaaaaattgt	atttgcttgg	1860
atttatattgc	ctctcaataa	agcttttgac	ttggaaaaaa	aaaaaaa		1907

**THIS PAGE BLANK (USPTO)**

<210> 36  
<211> 1839  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 3143411CB1

<400> 36  
gcggtccgtc ggtggcctag agatgctgct gccgcggttg cagttgtcgc gcacgcctct 60  
gcccgccagc ccgctccacc gccgtagcgc ccgagtgtcg gggggcgcac ccgagtcggg 120  
ccatgaggcc gggaaccgag ctacaggccg tgctgtctggc cgtgtctgctg gtggggctgc 180  
gggcccgcgac gggtcgcctg ctgagtgggc agccagtctg ccggggaggag acacagaggc 240  
cttgttataa agtcatttac ttccatgata cttctcgaag actgaacttt gaggaagcca 300  
aagaagcctt caggagggat ggaggccagc tagtcagcat cgagtctgaa gatgaacaga 360  
aactgataga aaagttcatt gaaaacctct tgccatctga tgggtgacttc tggattgggc 420  
tcaggaggcg tgaggagaaa caaagcaata gcacagcctg ccaggacctt tatgcttggg 480  
ctgatggcag catatcacaa tttaggaact ggtatgtgga tgagccgtcc tggggcagcg 540  
aggctctcgt ggtcatgtac catcagccat cggcaccgcg tggcatcgga ggcccctaca 600  
tgttccagtg gaatgatgac cgggtgcaaca tgaagaacaa tttcatttgc aaatatcttg 660  
atgagaaacc agcagttcct tctagagaag ctgaagggtga ggaaacagag ctgacaacac 720  
ctgtacttcc agaagaaaca cagggaagaag atgccaaaaa aacattttaa gaaagtagag 780  
aagctgcctt gaattctggc tacatcctaa tccccagcat tccccctctc ctccctcttg 840  
tggtcaccac agttgtatgt tgggttttggg tctgtagaaa aagaaaacgg gagcagccag 900  
accctagcac aaagaagcaa cacaccatct ggccctctcc tcaccaggga aacagcccg 960  
acctagaggc ctacaatgtc ataagaaaac aaagcgaagc tgacttagct gagaccgggc 1020  
cagacctgaa caaatattca ttccgagtgt gttcgggaga agccactccc gatgacatgt 1080  
cttgtgacta tgacaacatg gctgtgaacc catcagaaag tgggtttgtg actctggtga 1140  
gcgtggagag tggatttgtg accaatgaca tttatgagtt ctcccagac caaatgggga 1200  
ggagtaagga gtctggatgg gtggaaaaatg aaatatatgg ttattaggac atataaaaaa 1260  
ctgaaactga caacaatgga aaagaaatga taagcaaaat cctcttattt tctataagga 1320  
aaatacacag aagggtctatg aacaagctta gatcagggtcc tgtggatgag catgtggtcc 1380  
ccacgacctc ctgttggacc cccacgtttt ggctgtatcc tttatcccag ccagtcaccc 1440  
agctcgacct tatgagaagg taccttggcc aggtctggca catagtagag tctcaataaa 1500  
tgtcacttgg ttggttgtat ctaactttta agggacagag cttacctggc cagtataaaa 1560  
gatgggctgt ggagcttggg aaaccacctc tgttttccct gctctataca gcagcacata 1620  
ttatcataca gacagaaaat ccagaatctt ttcaaagccc acatatggta gcacaggctg 1680  
gcctgtgcat cggcaattct catatctgtt tttttcaaag aataaaatca aataaagagc 1740  
aggaaacaga aaaaaaaagg agatagggtg gtagatgatg gtaaggagtg cttaaagagg 1800  
tgtgggaaga gttgattgaa tggagggggg cccttgttt 1839

<210> 37  
<211> 503  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 3170835CB1

<400> 37  
cggctcgaga tgaaattcgc catcgctcctg ttgcgccctct ttgcggtggc cctggctgcc 60  
cctactgtcg aggtcctgag atcggatagc aatgttggaa tcgataacta ctcatatgca 120  
ggtgaaacca gcgacggtac ctogaagagc gaggagggtg tgctgaagaa cgccggcacc 180  
gagctagagg ccattctaac ccacggctcc ttcagctacg tgggccctga tggccagacc 240  
tacaccgtca cctacgtggc cgatgagaac ggattccagc cccagggtgc tcatctgccc 300  
gttgcctccg ttgcctaagc taaggagtgt aaataaattg tttataaaag caaaaaaact 360  
ggcagaaaga acccaaaagt cgtagacctt catgctagag agaattgtat gtctgccgaa 420  
gacaaccgta cccgctcgag caatcttacg gccactagtg gatcctgcgc atgaggacac 480  
cgagagcgag cgtggggggg ggc 503

<210> 38  
<211> 2154  
<212> DNA  
<213> Homo sapiens

**THIS PAGE BLANK (USPTO)**



<220>  
<221> misc\_feature  
<223> Incyte ID No: 3550808CB1

<400> 38  
gaggacagga tgaggccccg cctctcattt ctcttagccc ttctgttctt ccttggccaa 60  
gctgcagggg atttggggga tgtgggacct ccaattccca gccccggctt cagctctttc 120  
ccaggtgttg actccagctc cagcttcagc tccagctcca ggctcgggctc cagctccagc 180  
cgcagcttag gcagcggagg ttctgtgtcc cagttgtttt ccaatttcac cggctccgtg 240  
gatgaccgtg ggacctgcca gtgctctgtt tccctgccag acaccacctt tcccgtggac 300  
agagtggaaac gcttggaaatt cacagctcat gttctttctc agaagtgtga gaaagaactt 360  
tctaaagtga gggaatatgt ccaatttaatt agtgtgtatg aaaagaaact gttaaaccta 420  
actgtccgaa ttgacatcat ggagaaggat accatttctt aactgaact ggacttcgag 480  
ctgatcaagg tagaagtga ggagatggaa aaactggta tacagctgaa ggagagtttt 540  
ggtggaaagt cagaaaattgt tgaccagctg gaggtggaga taagaaatat gactctcttg 600  
gtagtaggag ttgacagact agacaaaaac aatgtccttg ccattcgccg agaaatcggtg 660  
gctctgaaga ccaagctgaa agagtgtgag gcctctaaag atcaaaacac ccctgtcgtc 720  
caccctcctc ccactccagg gagctgtggt catggtgggtg tgggtgaacat cagcaaacccg 780  
tctgtggttc agctcaactg gagaggggtt tcttatctat atggtgcttg gggtagggat 840  
tactctcccc agcatccaaa caaaggactg tatgggtgg cgccattgaa tacagatggg 900  
agactgttgg agtattatag actgtacaac acactggatg atttgcatt gtatataaat 960  
gctcgagagt tgcggatcac ctatggccaa ggtagtggtg cagcagttta caacaacaac 1020  
atgtacgtca acatgtacaa caccgggaat attgccagag ttaacctgac caccaacacg 1080  
attgctgtga ctcaaacctc ccctaattgt gcctataata accgcttttc atatgctaatt 1140  
gttgcttggc aagatattga ctttgcgtgt gatgagaatg gattgtgggt tattttattca 1200  
actgaagcca gcaactggtaa catggtgatt agtaaactca atgacaccac acttcaggtg 1260  
ctaaacactt ggtataccaa gcagtataaa ccatctgctt ctaacgcctt catggtatgt 1320  
gggttctgt atgccacccg tactatgaac accagaacag aagagatttt ttactattat 1380  
gacacaaaac cagggaaaga gggcaaaacta gacattgtaa tgcataagat gcaggaaaaa 1440  
gtgcagagca ttaactataa cccttttgac cagaaaactt atgtctataa cgatggttac 1500  
cttctgaatt atgatctttc tgtcttgacg aagccccagt aagctgttta ggagttaggg 1560  
tgaaagagaa aatgtttgtt gaaaaaatag tcttctccac ttacttagat atctgcaggg 1620  
gtgtctaaaa gtgtgttcat tttgcagcaa tgttttaggt catagtctta ccacactaga 1680  
gatctaggac atttgtcttg atttgggtgag ttctcttggg aatcatctgc ctcttcaggc 1740  
gcattttgca ataaagtctg tctaggggtg gattgtcaga ggtctagggg cactgtgggc 1800  
ctagtgaagc ctactgtgag gaggttcac tagaagcctt aaattaggaa ttaaggaaact 1860  
taaaactcag tatggcgtct agggattctt tgtacaggaa atattgcca atgactagtc 1920  
ccatccatg tagcaccact aattcttcca tgcttggaa aaacctgggg acttagttag 1980  
gtagattaat atctggagct cctcgaggga ccaaactccc aactttttt tcccctcact 2040  
agcaccctga atgatgcttt gtatgtggca gataagtaaa tttggcatgc ttatatattc 2100  
tacatcctga aagtggcctg gttttatgga gaagaggcct ctttatgcag tcaa 2154

<210> 39  
<211> 733  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 3683905CB1

<400> 39  
gcgagcaggt tgccgagcgc gccccgtccc tcgcgcgcga tgctccctg gacggcgcctc 60  
ggcctggccc tgagcttgcg gctggcgctg gcgcggagcg gcgcgggagc gcggtccacc 120  
agcatcagcc ccccgagggg acctgatgtt cctgctggac agctcagcca gcgtctctca 180  
ctacgagttc tcccgggttc gggagtttgt ggggcagctg gtggctccac tgcccctggg 240  
caccggggcc ctgcgtgcca gtctggtgca cgtgggcagt cggccataca ccgagttccc 300  
cttcggccag cacagctcgg gtgaggctgc ccaggatgcg gtgcgtgctt ctgcccagcg 360  
catgggtgac acccacactg gcctggcgct ggtctatgcc aaggaaacagc tgtttgctga 420  
agcatcaggt gcccgccag ggtggccaa agtgcgtgtg tgggtgacag atggcgctc 480  
cagcgaccct gtggggcccc ccattgcagga gctcaaggac ctgggcgtca ccgtgttcat 540  
tgtcagcacc ggccgagga acttcctgga gctgtcagcc gtgcctcag cccctgccga 600  
gaagcacctg cactttgtgg acgtggatga cctgcacatc attgtccaag agctgagggg 660  
ctccattctc gacgcgatgc ggcgcgaagc ttattccctt tagtgagggt taattttagc 720  
ttgcactggc cgt 733

<210> 40

**THIS PAGE BLANK (USPTO)**



<211> 718  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 4062841CB1

<400> 40  
gccggcgcca gggcaggcgg gcggctggca gctgtggcgc cgacatggct gcgctgggtg 60  
agccgctggg gctggagcgg gacgtgtccc gggcggttga gctcctcgag cggctccagc 120  
gcagcgggga gctgccggcg cagaagctgc aggccctcca gcgagttctg cagagccgct 180  
tctgtctcgc tatccgagag gtgtatgagc agctttatga cacgctggac atcacgggca 240  
gcgccgagat ccgagcccat gccacagcca aggccacagt ggctgccttc acagccagcg 300  
agggccacgc acatcccagg gtagtggagc tacccaagac ggatgagggc ctaggcttca 360  
acatcatggg tggcaagag caaaactcgc ccattctacat ctcccgggtc atcccagggg 420  
gtgtggctga ccgccatgga ggcctcaagc gtggggatca actgttgctg gtgaacgggtg 480  
tgagcgttga gggtagcag catgagaagg cggtaggagc gctgaaggcg gccaggggct 540  
cggtagagct ggttgtccgt tacacaccgc gagtgtgga ggagatggag gcccggttcg 600  
agaagatgct ctctgccgc cggcgccaac agcatcagag ctactcgtcc ttggagtctc 660  
gaggttgaaa ccacagatct ggacgttcac gtgcactctc ttctgtaca gtatttat 718

<210> 41  
<211> 1235  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 6394358CB1

<400> 41  
gccgccggag ccgccgccgc agcggggagc gggagcccc gggggccccg ccaccgccgc 60  
cgcccgccgt cacctaccgc gactggatcg gccagagtta ctccgagggtg atgagcctca 120  
acgagcactc catgcaggcg ctgtcctggc gcaagctcta cttgagccgc gccaaagctta 180  
aagcctccag ccggacctcg gctctgtctt ccggcttcgc catggtggca atggtggagg 240  
tgagctggga cgctgaccac gactaccac ccgggctgct catcgccctc agtgcttga 300  
ccacagtgtg ggtggctgtg cacctgtttg cgctcatgat cagcacctgc atcctgcca 360  
acatcgaggc ggtgagcaac gtgcacaatc tcaactcggg caaggagtcc ccccatgagc 420  
gcatgcaccg ccacatcgag ctggcctggg ccttctccac cgtcatcggc acgctgtctc 480  
tcctagtga ggtgggtgtg ctctgtctgg tcaagttctt gccctcaag aagcagccag 540  
gccagccaag gccaccagc aagcccccg ccagtggcgc agcagccaac gtcagacca 600  
gcggcatcac cccgggccag gcagctgcca tcgctcgac caccatcatg gtgcccttcg 660  
gcctgatctt tatcgtcttc gccgtccact tctaccgctc actggttagc cataagactg 720  
accgacagtt ccaggagctc aacgagctgg ccgagtttgc ccgcttacag gaccagctgg 780  
accacagagg ggaccacccc ctgacgccc gcagccacta tgcttaggcc catgtgtgtc 840  
gggcccttcc agtgcttttg ccttacgccc ttcccttga ccttgtcctg cccagcctc 900  
acggacagcc tgcgcagggg gctgggcttc agcaaggggc agagcgtgga gggaagagga 960  
ttttataag agaaatttct gcactttgaa actgtcctct aagagaataa gcatttctg 1020  
ttctccagc tccaggtcca cctcctgttg ggaggcgggtg gggggccaaa gtggggccac 1080  
acactcgtg tgtccctct cctcccctgt gccagtggca cctgggtgcc tcctcctgtc 1140  
ctgtccgtct caacctccct cccgtccagc attgagtgtg tacatgtgtg tgtgacacat 1200  
aaatatactc ataaggacac ctccaaaaaa aaaaa 1235

<210> 42  
<211> 11648  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 2847752CB1

<400> 42  
atggcgaggc ggccgccgtg gcggggcctc gggggacggt cgacccccat actcctgtc 60  
cttctcctct ctttgttccc cctcagccag gaggagctgg gggcggtgg gcaccagggc 120  
tgggaccag gcttagctgc cactacgggg ccaagggcgc atatcggtgg cggagcctta 180

**THIS PAGE BLANK (USPTO)**

gctctttgtc	cggagtcttc	cggggtccgg	gaggatgggg	ggcctggcct	gggggtcagg	240
gagcctatct	tcgtgggggt	ccgagggaga	aggcaaagcg	cccggaaatag	tcgagggccc	300
cctgagcagc	cgaatgagga	gctggggatt	gaacacggcg	tccagccatt	gggagccgc	360
gaacgagaga	caggacaggg	accagggtct	gtgttatact	ggcgcccaga	ggtctcctct	420
tgccggcgga	caggaccttt	gcaaagaggt	agtctgtcac	caggggctct	gtcctcaggg	480
gtcccgggct	cggggaacag	ctcgcccctc	ccttcagact	ttttgattcg	gcaccacggg	540
cccaagccgg	tgctctccca	gcggaacgct	gggacaggct	cccgaaaag	agtgggcacc	600
gcgcgctgct	gtgggggaatt	atgggcaaca	gggagcaagg	gtcagggcga	gagagccacg	660
acatccggag	cagaaaggac	agcccccg	cggaactgtc	ttccaggggc	ctcgggatct	720
ggccccgagc	tggattcagc	accacgcacg	gcgaggacag	ctcctgcac	aggttcagca	780
ccccgcgagt	ctcggacagc	tcccgagccg	gcgcccagc	gcatgcgctc	ccggggctct	840
ttccgctgcc	gcttctctcc	gcagcgcccc	gggcgcgctc	ccccgggact	ccccggcctg	900
cctgaagcca	ggaaagtaac	ctcggcgaac	cgggcacgct	ttcgtcgcgc	cgcaaacgcg	960
caccgcgagt	ttccgcagta	caactaccag	acgctgggtg	cggagaatga	ggcagcaggc	1020
accgcgggtg	tacgcgtggg	tgctcaggac	ccggacgccc	gcgaggccgg	gcgcctagtc	1080
tactcgctgg	cggcactcat	gaacagccgc	tcgctggagc	tggttcagcat	cgaccgcagc	1140
agcgcgctta	tcctgtacgg	ggcagctctg	gcgcgcgaga	gcatggagcg	tcactacctg	1200
cgtgtgaccg	cgcaggacca	cggtctcgcc	cgctctctcg	ccaccacgat	ggtggccgtg	1260
acagttagccg	accgcaacga	ccactcgccg	gttttttgag	aagcgcagta	ccgggagacc	1320
cttcgcgaga	atgtggagga	gggctacctc	atcctgcagc	tgctgtgccac	tgacggcgac	1380
gcgcccccca	acgccaacct	gcgctaccgc	ttcgtggggc	cgccagctgc	gcgcctgcga	1440
gctgcgcgcg	ccttcgagat	tgatccacgc	tcgggctcca	tcagcaccag	cgcccgagtg	1500
gaccgcgagc	acatggaaag	ctatgagctg	gtggtggaag	ccagcgacca	gggcccagaa	1560
ccccggccgc	gctcggccac	tgctgcgcga	cacataactg	tgctagacga	gaacgacaat	1620
gctctctagt	tcagcgagaa	gcgctacgtg	gcgaggtgct	gcgaggtatg	gcgccccac	1680
acagtctgtg	tgccgcgtcac	ggccactgac	cgggacaagg	acgccaacgg	attggtgcac	1740
tacaacatca	tcagtggcaa	tagccgtgga	cacttttgcca	tcgacagcct	cactggcgag	1800
atccagggtg	tgccacctct	ggactctcag	gcagagagag	agtatgcctt	gcgcctcagg	1860
gcgcaggatg	ctggccggcc	accgctgtcc	aacaacacgg	gcctggccag	catccaggtg	1920
gtggacatca	atgaccacat	tcttattttt	gtcagcacgc	ccttccaagt	ttctgtcttg	1980
gaaaatgtct	ccttgggtca	ctcagtcac	cacattcagg	cagtcgatgc	agaccatggg	2040
gagaatgccca	gattggagta	ctccctaact	gggtgtggcc	ctgatactcc	ttttgtgata	2100
aacagcgcca	ctggctgggt	ctctgtgagt	ggccccctgg	accgtgagtc	tgtgtggcat	2160
tacttctttg	gtgtggaggg	tcgagaccat	ggctcacc	cactctctgc	ctcagccagt	2220
gtcaccgtga	ctgtgctgga	cgttaatgac	aatcgccctg	agttcacaat	gaaggagtac	2280
cacctacgac	tgaatgagga	tgacgtgtg	ggcaccagtg	tggtcagcgt	gaccgcagta	2340
gaccgtgatg	ccaacagtgc	catcagctac	catcacacag	gcggcaacac	ccggaaatgc	2400
tttgccatca	gcacccaggg	gggtgtgggt	ctggtgactc	tggtctctgc	actggactac	2460
aagcaggaac	gctacttcaa	gctggtacta	actgcacttg	accgtgccct	tcagtatcac	2520
tgctatgtgc	acatcaacat	cacagatgcc	aacactcctc	ggccgggtctt	tcaaagtgc	2580
cactactcag	tgagtgtgaa	tgaagatcgg	ccaatgggta	gcaccatagt	ggtcatcagt	2640
gcctctgatg	atgacgtggg	tgagaatgct	cgtatcacct	atctcctgga	ggacaacctg	2700
ccccagttcc	gcattgatgc	agactcagga	gccattacat	tacaggcccc	attagactat	2760
gaggaccagg	tgacctacac	cctggctatc	acagctcggg	acaatggcat	cccacagaag	2820
gcagacacta	cttatgtgga	gggtgatggc	aatgacgtga	atgacaatgc	tccacaattt	2880
gtggcctccc	actatacagg	gctggtctct	gaggatgccc	cacctttcac	cagtgtcctg	2940
cagatctcag	ccactgaccg	ggatgctcat	gccaatggcc	gggtccagta	cactttccag	3000
aatgggtgaag	atggggatgg	agattttacc	attgagccca	cctctggaat	tgtccgtaca	3060
gtaaggcggc	tagaccggga	ggcagtatca	gtgtatgagt	tgactgccta	cgcagtggac	3120
agaggtgtgc	ccccactccg	gactccagtc	agtatccagg	tgatgggtgca	ggatgtgaac	3180
gacaatgcac	ctgtcttccc	agctgaggag	tttgagggtg	gggtgaaaga	gaatagcatt	3240
gtgggctcag	tggtggccca	gatcactgca	gtggaccctg	acgaaggccc	caatgcccc	3300
ataatgtacc	agatcgtgga	ggggaacatc	ctgagctgtg	tccaaatgga	catcttctct	3360
ggagaactga	cggcactcat	tgacctagac	tatgaggctc	gccagaata	tgtgattgtg	3420
gtgcaggcca	catctgctcc	tttggtcagc	cgggccactg	tgacgctccg	cctgggtgac	3480
cagaatgaca	acagccctgt	gctcaacaac	ttccagatcc	tcttcaacaa	ctatgtatcc	3540
aaccgttcag	acaccttccc	gtcgggcatt	attgggcgca	ttccagctta	tgaccccgat	3600
gtctccgacc	acctcttcta	ctcctttgag	cgtggcaatg	agctgcagct	gctggtagtc	3660
aaccagacca	gtggggagct	gcgactcagc	cgaaagctag	acaataaccg	cccactgggtg	3720
gcctccatgt	tggtgactgt	cacagatggc	ctgcacagcg	tgacggcgca	gtgtgtgctg	3780
cgctgggtca	tcatacggga	ggagtgtctg	gccaacagcg	tgaccgtgcg	ccttgagaac	3840
atgtggcagg	agcgcttcc	gtcaccgctg	ctgggcccgt	tctcagagg	cgtggctgcg	3900
gtgctcgcta	cgcccgctga	ggacgtcttc	atcttcaaca	tccagaacga	cacagacgta	3960
gggggcacccg	tgctcaatgt	gagtttctct	gcgctagctc	cacgtggggc	cgggggcggc	4020
gctgcagggc	ctcgtttcag	ctccgaggag	ctgcaagagc	agttgtacgt	gcgcggggcg	4080
gcgctggcgg	ctcgctccct	gctcgacgta	ctgcccctcg	acgacaacgt	gtgcctgcga	4140
gagccctgtg	agaactacat	gaaatgcgtg	tccgtgctcc	gctttgactc	gtccgcgccc	4200

**THIS PAGE BLANK (USPTO)**

ttcctggcct	cggcctccac	gctgtttccga	cccatccagc	ccatcgctgg	cctgcgctgc	4260
cgctgcccgc	cgggattcac	gggagacttt	tgcgagaccg	agctcgacct	ctgctactcc	4320
aacctcagtc	gcaacggcgg	agcctgcgcg	cggcgcgagg	gaggctacac	gtgcgtctgc	4380
cgccccgcgt	tcaccggaga	ggactgcgag	ctggacaccg	aggccggccg	ctgcgtgccg	4440
ggcgtctgcc	gcaacggggg	cacctgcacc	gacgcgcccc	acggcggtct	tcgctgccag	4500
tgccccgcag	gcggcgcctt	cgagggcccc	cgctgcgagg	tggctgcgcg	ctccttcccc	4560
cccagttcgt	tcgtcatggt	tcgcggcctg	cggcagcgat	tccaccttac	gctgtccctc	4620
tcgttcgcga	cagtgcagca	gagcgggctg	ctcttctaca	acgggcgcct	gaacgagaag	4680
cacgacttcc	tggccctgga	actcgtggct	ggccaagtgc	ggctcacata	ttccacgggt	4740
gaatccaaca	ccgtggtcag	cccacagtt	ccagggggct	tgagtgcagg	gcaatggcat	4800
acagtgcac	tgagatacta	caacaagccc	cggacagatg	ccctaggggg	tgcaacgggc	4860
ccctccaagc	acaagtggtc	tggtctaagc	gtgatgtggc	gtgatgtggc	cgtaggggct	4920
cagtttggtg	ctgagattgg	caactactca	tgcgcggtcg	ctggtgtgca	aacaagctcc	4980
aagaagtcct	tggacctgac	gggcctctct	cttctgggag	gtgtccccaa	cctccccgag	5040
aactcccccg	tatcccataa	ggacttcact	ggctgtatgc	gggacctgca	cattgatggc	5100
cgccgggtgg	acatgcgggc	ttttgtcgca	aataatggca	ccatggcagg	ctgccaagcc	5160
aagctacact	tttgtgactc	aggccctgc	aagaacagtg	gcttctgctc	ggagcgctgg	5220
ggcagcttca	gctgcgactg	ccctgtgggc	ttcgcgggca	aagactgtca	gcttactatg	5280
gccccctccc	accattttccg	tggcaacggc	acactgagct	ggaactttgg	aagtgcacatg	5340
gctgtgtctg	tgccatggta	cctggggctg	acatgtcgga	cacgggcaac	gcaggggggtc	5400
ctgatgcaag	tgcaaggctg	gccacacagc	acgctccttt	gccagctaga	tcgggggtta	5460
ctgtctgtga	cagtgcaccag	gggctcgggc	cggtctccc	atctccttct	ggaccaggtg	5520
actgtcagtg	atggccgggtg	gcacgatctg	cggtctggag	tgcaaggagga	accaggtggc	5580
cgccggggcc	accatgtcct	tatggtctca	ctggagctta	gcctcttcca	ggacacatgt	5640
gcggtgggga	gtgagctgca	gggcctgaag	gtaaagcagc	tccacgtggg	aggcctgccc	5700
cccggcagtg	cagaggaggc	tcctcagggt	ctgggttggct	gcacccagcc	accgagtga	5760
tgccggacctg	gctgtgttgt	gaccaacgcc	tgtgcctctg	ggccctgccc	acctcacgca	5820
gactgcgggg	acctctggca	gaccttttct	tgcaactgcc	agccagggtta	ctacggccca	5880
ggctgtgtgg	atgcctgcct	cctgaacccc	tgctcagaacc	agggatcatg	ccggcacctg	5940
ccaggagccc	cccatggcta	tacctgtgac	tgtgtgggtg	gctatttcgg	gcaccactgt	6000
gagcacagga	tggaccagca	gtgcccacgg	ggctgggtgg	ggagcccaac	ctgtggcccc	6060
tgcaactgtg	atgttcacaa	aggttttgat	cccaactgca	acaagacaaa	tgggcagtgt	6120
cactgcaagg	agttccacta	ccgaccgcgg	ggcagtgcct	cttgccctcc	atgtgactgc	6180
tacctgtgtg	gtccacacct	gcgctcatgt	gcacccacac	gcgggcagtg	ccctgtctgc	6240
ccaggagccc	ttggcccgca	gtgcaacagc	tgtgacagtc	ccttcgcaga	ggtgacagcc	6300
agcggctggc	gggtgcctca	tgatgcctgc	ctgagctccc	tgagatctgg	tggtgtgtgg	6360
ccccagacaa	agtttgccgt	cctggccaca	gtgccctgtc	cccggggggc	cctgggattg	6420
cggggtgcag	gtgctgtctg	gcggctgtgt	gatgaggccc	aggggtggct	ggagcccagc	6480
ctcttcaact	gtacctcccc	tgcccttcca	gagctcagtc	tgctgtcggg	tgccctagag	6540
ctgaacaaga	cggcatggga	taccatggag	gccaagaagc	tggctcagcg	gctacgggag	6600
gtgactggcc	acactgacca	ctatttttagc	caagatgttc	gagtcactgc	ccgcctgtctg	6660
gcccacctgc	tggccttcga	gagccatcag	cagggcttcg	ggctgacagc	cacacaggat	6720
gcccacttca	atgagaatct	gctgtggggc	ggctctgcac	tgcttgcccc	agagacaggg	6780
gacttgttgg	cggcgctggg	gcagcggggc	cctgggggct	ccccaggcag	cgctgggactg	6840
gtgaggcacc	tggaggagta	tgcagccaca	ctcgcaaggga	atatggaact	cacataacctg	6900
aatcccatgg	ggctgggtgac	gcctaataatc	atgctcagca	ttgaccgcat	ggagcacccc	6960
agttctcccc	ggggggcccc	tcgctaccct	cgctaccata	gcaacctctt	tcgaggccag	7020
gatgctctgg	atcctcacac	ccatgtgctg	ctgcctcccc	agtcgccacg	gccatcccca	7080
tctgaagtcc	tgccacaaag	cagcagcata	gaaaactcca	ccacctcaag	tgtggtcccc	7140
ccaccagccc	cgccagagcc	agagcctggg	atctccatta	tcattctcct	cgtttaccgc	7200
accttagggg	gactgtctcc	tgcccagttc	caggcagaac	gcccagggtgc	caggcttctct	7260
cagaacaagg	tcattgaactc	cccgttggtc	agcgtggctg	tgctccacgg	acgcaacttc	7320
ctaaggggaa	tcctggagtc	cccatcagc	ctagagtttc	gcctgtctaca	gacagcgaa	7380
cggagcaagg	cgatctgtgt	gcagtgggac	ccacctggcc	tggcggagca	gcatgggtgtg	7440
tggacagcac	gggactgcga	gctggtgcac	aggaatgggt	cccacgcacg	gtgtcgcctgc	7500
agccggacag	ggacctttgg	ggtcctcatg	gatgcctctc	cccgtgagag	gctggagggc	7560
gacctggagc	tgctggctgt	gttcacccac	gtggctcgtg	ctgtgtctgt	ggctgcgctg	7620
gtgctgactg	cagccatcct	gctgagcctg	cgcagcctca	agtccaatgt	gcgtgggagc	7680
catgccaatg	tggcagccgc	cctgggggtg	gcagagctcc	tcttctctgt	ggggattcac	7740
aggaccacac	atcagctggg	gtgcactgca	ctgcgcctcc	tcctgcacta	cttcttctcc	7800
agcaccttcg	cgtggctctt	cgtgcagggg	ctgcacctct	accgcatgca	ggttgagcca	7860
cgcaacgtgg	accgcggcgc	catgcgcttc	taccatgccc	tgggctgggg	cgctccctgt	7920
gtgctgtctg	gccttgctgt	gggcctggac	cctgaggggt	atgggaaccc	tgacttctgc	7980
tggatctcag	tccacgagcc	ctcatctgg	agcttctctg	gccctgttgt	cctggtcata	8040
gtgatgaacg	ggacctgtgt	tctcctcgct	gcccgcacat	cctgtctcac	agggcagagg	8100
gaggccaaga	agacctctgc	actgacctct	cgcagctcct	tcctgtctgt	tctgtctgtc	8160
agtgcctcct	ggctcttttg	gctcctggca	gtcaaccaca	gcacccctagc	cttccactac	8220

**THIS PAGE BLANK (USPTO)**



ctccatgctg	gactctgcgg	cctccagggc	ctggcggtgc	tgtgtctctt	ctgtgtccta	8280
aatgcagatg	ctcgggctgc	ctggatgcca	gcctgtctgg	gcaggaaggc	agcgccctgag	8340
gaggcaaggc	cagcacctgg	gctgggacct	ggggcctaca	acaacacggc	tctcttttgag	8400
gagagtggcc	tcctccgcat	cactctgggc	gcctccaccg	tctcctctgt	gagcagtgcc	8460
cgctccggcc	ggaccagga	ccaggacagc	cagcggggcc	gcagctacct	cagggacaat	8520
gtcctgggtc	gacatggctc	agccgctgac	cacactgacc	acagcctcca	ggctcatgct	8580
ggccccactg	acctggacgt	ggccatgttc	catcgagatg	ctggcgcgaga	ctccgactct	8640
gacagtgacc	tgtccttggg	ggaggagagg	agtctctcca	ttccatcttc	agaaagcgag	8700
gacaatggcc	ggacgcgggg	gcgcgtccaa	cgccactct	gccgagcagc	ccagagtggg	8760
aggctcctca	cccaccccaa	agatgtggat	ggcaatgacc	tctgttcta	ctggccagcc	8820
ctgggggagt	gcgaggcagc	cccctgtgct	ctgcagactt	ggggctctga	aaggcgccctg	8880
gggctggaga	ccagcaagga	tgcagctaac	aacaaccage	cagaccgggc	cctgaccagt	8940
ggggatgaga	cttctctggg	ccggggccag	cgccagagga	aaggcatcct	gaaggaccgg	9000
ttgcaatacc	cactggtgcc	acagaccgga	ggtgccccctg	agctgtcctg	gtgccgtgca	9060
gccaccttgg	gccaccgtgc	tgtgccagct	gcctcttacg	gtcgcatcta	tgtggcgggg	9120
ggcacgggca	gcctttcaca	gccagccagc	cgctactctt	ctagagaaca	gctggacctg	9180
ctcctccggc	ggcaactgag	ccgtgagcga	ctagagggaag	cccctgcccc	tgtttctaac	9240
cccctgagcc	ggccagggtc	ccaggaatgc	atggatgctg	caccaggccg	actggagccc	9300
aaagatcggg	gcagcacctt	gccacggagg	cagccacctc	gggactaccc	tggcgccatg	9360
gctggcgctc	tcgggtcacg	ggatgcgctc	gacttagggg	cacctcgaga	gtggttgagc	9420
acgtccgctc	cgcccggcgg	cacccgggac	cttgaccacac	agccccacc	tctgttccag	9480
tctccccagc	ggcaactctc	aagggaaccc	ctcttgccat	cccggccgct	ggactctctg	9540
tctaggagct	cgaactctcg	ggagcagctg	gaccagggtgc	ctagccggca	cccctcaccg	9600
gaagcccttg	ggccactccc	gcagctgctc	agagctaggg	aggactcggg	cagtggcccc	9660
agccatggcc	cctccacaga	acagttggac	attctttcct	ccatccttgc	ctctttcaac	9720
tctcggccc	tctcctctgt	gcaatcttca	agcacacctt	tgggcccctca	caccactgcc	9780
acacctttctg	ccacagcctc	tgtgcttggg	ccctccacgc	cacgttctgc	cacgtctcac	9840
agcatctcgg	agctgtcgcc	agactcagaa	gttcccagaa	gtgaggggtca	ctcctgaggg	9900
gatgacggcg	tggacagagga	acagctgagg	gtcagagggt	atctaggcta	acaggagaga	9960
ctccaggagt	gggggcagat	cccaaggcag	cctcctgctc	cccagtgggtg	ggtgccccag	10020
ctctacctgg	tgtggcaggg	ctgaggctcc	atgtgcatct	gtgagcatgc	gtgtgacagg	10080
tgcagagacg	ggggactgga	gggagacttt	tatacgtttt	gtacctttgt	aaccagagag	10140
atgcttatgt	tatttttcag	cttttctgtc	tctggggggg	tttgaggctg	ggctgggagg	10200
gggagggaga	tagagggaga	gatgcagttt	gaccccatct	gggtcctgag	caaaccctat	10260
gctcatctct	ctctccttcc	tgggggtggac	tcagatgggt	gggacacatg	ccttccctccc	10320
cctattccac	cccagaagtg	atctgagtat	cgtcaggggc	ccaaagtaca	gaattgttct	10380
ttgtttttta	ttgaatgtct	caaaggccaa	ctgtctgggg	ctggggggttg	gtcttggaaa	10440
caggggtcct	ctgacttcc	catgggggct	tgtcataacc	gccccctctg	gtggatgtgt	10500
gtgtttatta	tgtggagtc	ctgccactta	ctgccttatg	acctaggact	gatgctgtgg	10560
ggtgctgggt	gagcagctga	tgtcgtgttt	acagagcaag	gcttccctgt	ctcccacggg	10620
gaggggctcg	ggcctctagt	cagacattcc	tgcagagggt	cggtggaggg	gtcattcacc	10680
tgcccctgca	gcaagcaaaa	gttgtctgtg	gtgccatttg	attccctgac	actgccccct	10740
gcttgaattg	attccgaagg	gtagggtggg	aagggtgagca	aagggagcag	aaacaaggga	10800
attcaagacc	cagaatgtag	gtgccactgc	ctcctatgtt	tacaggatcc	tccgtggccc	10860
taggcacctg	ggctgcagga	agtgactccg	ttccactcct	cctttattcc	cttaaaaagg	10920
gaaaaatgac	tgttacgacc	ctgttcacaa	aactcttact	tttgctattt	tgtctgctgt	10980
ccagaactga	agactttaaa	attttgttac	tgtttacaag	tccagattca	aaaaatgttt	11040
ttactttggt	tacaactcaa	aactttgagt	tttacacttt	gtttacagta	gataattttt	11100
tttcccttgt	ttccaagtga	aaggtaggga	aagtgggaga	gggacttgga	ggaccacact	11160
gtgaggaccc	tgacctggcc	atcttgaggg	gttttctaac	ccccaggctc	cccaggccga	11220
aggtcagcct	tgagtcccgt	ttaacagcag	atccagaaga	ccttgagagt	aggcgctcct	11280
taaccacggg	ggagagtggc	tgtgcagggc	tggggggtgg	tctgtgcaga	cacctctca	11340
cccaccaccc	catgcatact	cttgggaagc	agcttccctg	gagattagaa	attctacttc	11400
cctgactgga	gctaaatccc	accagccagg	acccaaactc	tccttaccga	gaaggacccc	11460
agctcttgaa	gggctgagtg	gcctgtctgg	ggtgggaggg	tgtctttact	atgtcctagg	11520
tttcgtagat	gccctctct	gggggtcccc	tcctccagcc	cagcgccct	ctttcctgtc	11580
tgtgtaaatt	gttccgtgaa	gcgcgctct	gttttgggaa	taaacttcta	tagaaaacaa	11640
aaaaaaaa						11648

**THIS PAGE BLANK (USPTO)**